Formaldehyde column density measurements as a suitable pathway to estimate near-surface ozone tendencies from space

Jason R. Schroeder[1,2], James H. Crawford[1], Alan Fried[3], James Walega[3], Andrew Weinheimer[4], Armin Wisthaler[5,6], Markus Müller[5], Tomas Mikoviny[6], Gao Chen[1], Michael Shook[1], Donald R. Blake[7], Glenn Diskin[1], Mark Estes[8], Anne M. Thompson[9,10], Barry L. Lefer[11,*], Russell Long[12], Eric Mattson[13]

[1] NASA Langley Research Center, Hampton, VA, USA
[2] NASA postdoctoral program, NASA Langley Research Center, Hampton, VA, USA
[3] Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA
[4] National Center for Atmospheric Research, Boulder, CO, USA
[5] Institute of Ion Physics and Applied Physics, University of Innsbruck, Austria
[6] Department of Chemistry, University of Oslo, Oslo, Norway
[7] Department of Chemistry, University of California Irvine, Irvine, CA, USA
[8] Texas Commission on Environmental Quality, Austin, TX, USA
[9] Department of Meteorology, Penn State University, University Park, PA, USA
[10] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[11] Department of Earth & Atmospheric Science, University of Houston, Houston, TX, USA
[12] National Exposure Research Laboratory, US EPA, Research Triangle Park, NC, USA
[13] Colorado Department of Public Health and Environment, Denver, CO, USA

[*] Now at NASA Headquarters, Washington, DC, USA

Corresponding author: Jason Schroeder (Jason.r.schroeder@nasa.gov)

Key Points:

- A correlation between column CH_2O and near-surface O_3 measurements was observed and could be useful for monitoring air quality from space.
- The strength of the correlation between column CH_2O and near-surface O_3 is highest when there is temporal variability in hydrocarbon emissions.
- The O_3-CH_2O relationship appears to be strongest in regions where emissions of biogenic VOCs dominate the local hydrocarbon mix.

Abstract

In support of future satellite missions that aim to address the current shortcomings in measuring air quality from space, NASA's DISCOVER-AQ field campaign was designed to enable exploration of relationships between column measurements of trace species relevant to air quality at high spatial and temporal resolution. In the DISCOVER-AQ dataset, a modest correlation ($r^2 = 0.45$) between ozone (O_3) and formaldehyde (CH_2O) column densities was observed. Further analysis revealed regional variability in the O_3-CH_2O relationship, with Maryland having a strong relationship when data were viewed temporally, and Houston having a strong relationship when data were viewed spatially. These differences in regional behavior are attributed to differences in VOC emissions. In Maryland, biogenic VOCs were responsible for ~28% of CH_2O formation within the boundary layer column, causing CH_2O to, in general, increase monotonically throughout the day. In Houston, persistent anthropogenic emissions dominated the local hydrocarbon environment, and no discernable diurnal trend in CH_2O was observed. Box model simulations suggested that ambient CH_2O mixing ratios have a weak diurnal trend (± 20% throughout the day) due to photochemical effects, and that larger diurnal trends are associated with changes in hydrocarbon precursors. Finally, mathematical relationships were developed from first principles and were able to replicate the different behaviors seen in Maryland and Houston. While studies would be necessary to validate these results and determine the regional applicability of the O_3-CH_2O relationship, the results presented here provide compelling insight into the ability of future satellite missions to aid in monitoring near-surface air quality.

1 Introduction

Great improvements in air quality have been made throughout the US over the past few decades, but progress has begun to plateau, leaving many regions with criteria pollutant levels that regularly exceed the national ambient air quality standards (NAAQS). In addition, the US Environmental Protection Agency (EPA) has recently lowered the exceedance level for ozone (O_3) from 75 parts per billion (ppbv) to 70 ppbv over an eight hour period [*US EPA*, 2015]. As a result, some regions that were previously in compliance with the NAAQS may now be considered non-attainment areas, and some previous non-attainment areas are expected to be re-classified with a worse ranking [*Downey et al.*, 2015]. Some of the difficulty in making further improvements in air quality can be attributed to increases in background levels of criteria pollutants and their chemical precursors [*Hudman et al.*, 2004; *Oltmans et al.*, 2006; *Lin et al.*, 2012a; *Simpson et al.*, 2012; *Cooper et al.*, 2014]. That is, long and medium-range transport (intercontinental and inter-regional transport) of airmasses may convolute our understanding of local air quality issues in non-attainment areas, thereby making it increasingly more challenging for scattered networks of surface monitoring stations to accurately depict regional air quality [*Paoletti et al.*, 2014; *Zhang et al.*, 2014; *Zoogman et al.*, 2014]. While satellite-based measurements of criteria pollutants could theoretically help fill in the gaps between surface monitoring sites, extracting near-surface concentrations of pollutants from column density measurements (such as those that are provided from satellites) has proven difficult [*Reed et al.*, 2013; *Flynn et al.*, 2014]. In this work, data collected during NASA's Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) field campaign is used to explore the possibility of using space-based measurements of formaldehyde (CH_2O) as a means to estimate near-surface photochemical conditions and understand its relationship with O_3. Given the difficulty in detecting near-surface

O_3 directly from space, any corroborating information on O_3 behavior and related exposure to poor air quality would raise the confidence in satellite information and its impact on decision making.

There are many challenges to using space-based platforms to probe surface air quality. Most notably, the vertical distributions of trace species provide inherent challenges in extracting near-surface measurements (and thus, relevant measurements for human exposure) from the total atmospheric columns measured by satellites [*Martin*, 2008; *Ichoku et al.*, 2012; *National Science and Technology Council*, 2013; *Streets et al.*, 2013; *Duncan et al.*, 2014]. Because the majority (~90%) of column O_3 is contained in the stratosphere, much research has focused on developing satellite retrieval algorithms that separate the tropospheric O_3 column from the stratospheric O_3 column [*Fishman et al.*, 2008]. However, under typical midlatitude conditions, O_3 in the planetary boundary layer (PBL, the mixed layer that extends 1-3 km above earth's surface) can contribute anywhere from 10 to 50% to the total tropospheric column, and substantial variability in upper tropospheric O_3 can make it difficult to attribute changes in the total tropospheric column to changes in PBL O_3 [*Martins et al.*, 2013; *Thompson et al.*, 2014]. Model simulations have suggested that multi-spectral retrieval techniques have the potential to derive a lower tropospheric or boundary layer O_3 column density, but these techniques cannot be implemented using the current constellation of air quality-relevant satellites [*Natraj et al.*, 2011; *Zoogman et al.*, 2011; *Hache et al.*, 2014].

Nitrogen Dioxide (NO_2) plays an important role in tropospheric O_3 formation, and is also routinely measured by satellite-based platforms. Column NO_2 also has a strong stratospheric contribution, however, its vertical profile is minimized in the upper troposphere and is heavily weighted towards the surface. Over polluted areas, near-surface NO_2 dominates the spatial variability in total column abundance. NO_2 also has substantial horizontal variability on sub-grid scales (i.e. smaller than satellite horizontal resolution), meaning that satellite measurements cannot fully resolve valuable details of the true NO_2 distribution. This also complicates validation efforts, as measurements from a discrete ground site are often inadequate to represent the average NO_2 abundance over a satellite footprint. In contrast to O_3 and NO_2, CH_2O has negligible presence in the stratosphere, and a tropospheric vertical profile that is heavily weighted towards earth's surface. In comparison to NO_2, it tends to have relatively low horizontal variability on sub-grid scales owing to its secondary source from hydrocarbon oxidation, which in contrast to highly-localized point sources (i.e. NO_2 emissions) tend to be spatially broader [*Fried et al.*, 2008; *Junkermann*, 2009; *Apel et al.*, 2012; *Baidar et al.*, 2013]. For this reason, satellite-based measurements of CH_2O may better represent actual conditions at the surface and be more easily interpreted to diagnose air quality conditions from space.

Previous work using satellite-based measurements of CH_2O have made inferences about near-surface hydrocarbon conditions and have been useful in deriving emissions estimates for biogenic VOCs over forested regions [*Millet et al.*, 2008; *Kefauver et al.*, 2014] and highly-reactive VOCs over urban areas [*Zhu et al.*, 2014]. In other work, co-located satellite measurements of CH_2O and NO_2 have been used to infer the regional sensitivity of tropospheric O_3 formation to changes in NO_x ($NO_x \equiv NO + NO_2$) or VOC reactivity (that is, whether regional O_3 production is NO_x or VOC-limited) [*Martin et al.*, 2004; *Duncan et al.*, 2010], and has proven useful for investigating near-surface photochemical environments in under-studied regions of the world [*Jin and Holloway*, 2015; *Mahajan et al.*, 2015]. While the CH_2O/NO_2 ratios described in these studies may be useful for observing near-surface photochemical environments, they do not

provide any information about the distribution of near-surface O_3 and therefore cannot be used to determine which regions would receive the greatest benefit from further regulation. Previous studies have noted a correlation between in-situ measurements of O_3 and CH_2O, but did not explore these behaviors in the vertical dimension and did not explore the possibility of using this correlation as a means to estimate near-surface O_3 tendencies from space [*Wert*, 2003; *Parrish et al.*, 2012]. Additional shortcomings of current satellite-based techniques arise from the nature of the satellite platforms themselves. Observations from polar-orbiting satellites are limited in their temporal resolution. For example, the ozone monitoring instrument (OMI), housed aboard NASA's AURA satellite, only collects measurements over a given region of earth once per day (usually around solar noon). While this is useful for looking at long-term averages of trace species, it is not useful for looking at the diurnal variations that may be important for understanding short-lived (i.e. a few days or less) air pollution events. Combining information from two satellites (e.g., OMI and SCIAMACHY or OMI and GOME-2) with different local overpass times have been used to assess diurnal variability of select trace gases, but the limited information from these two views still misses the late afternoon when air quality impacts are often most severe [*Boersma et al.*, 2008, 2009; *De Smedt et al.*, 2015].

A new collection of air quality satellites is on the horizon that will address the problem of temporal resolution. Making observations from geostationary orbit, these satellites will also deliver improved spatial resolution over key areas of the northern hemisphere. These include the Tropospheric Emissions: Monitoring of Pollution (TEMPO) mission over North America (http://science.nasa.gov/missions/tempo/), Sentinel-4 over Europe and North Africa (http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinels_-4_-5_and_-5P), and the Geostationary Environment Monitoring Spectrometer (GEMS) over East Asia (http://www.ballaerospace.com/page.jsp?page=319). The geostationary orbit of TEMPO, for example, allows for a temporal resolution of ~1 hour for O_3 and NO_2, and ~3 hours for CH_2O. Additionally, new multi-spectral retrieval algorithms are being optimized for these satellites to obtain information on boundary layer O_3 that would greatly improve our ability to monitor air pollution from space [*Zoogman et al.*, 2016]. While this multispectral technique is promising, additional indicators of near-surface O_3 behavior would be invaluable for validating the accuracy and robustness of techniques that directly retrieve near-surface O_3 from satellites.

In anticipation of these geostationary air quality observations, NASA conducted a series of field studies to examine the relationship between surface air quality and the vertical distribution of pollutants as they would be observed from space. The overall project, called DISCOVER-AQ, conducted four field studies in different parts of the U.S. (Maryland, California, Texas, and Colorado) at four different times (July 2011, January 2013, September 2013, and August 2014, respectively). The DISCOVER-AQ dataset enables the exploration of relationships between column amounts of O_3, NO_2, and CH_2O with very high spatial (both vertical and horizontal) and temporal (typically 3 measurements per day at each location) resolution, which effectively affords us a preview of the potential utility and applicability of future geostationary satellite missions. In the work presented below, aircraft data are used to investigate the relationship between O_3 and CH_2O, with an underlying goal of understanding whether column measurements of CH_2O might be useful for diagnosing near-surface O_3 behavior.

2 The Chemical Link Between O_3 and CH_2O

The impetus for examining the relationship between O_3 and CH_2O comes from their intertwined chemical origins. Both are secondary photochemical products influenced by human

emissions, with some intriguing examples of correlative behavior in previous literature examining their in situ behavior in Houston [*Wert*, 2003; *Parrish et al.*, 2012]. While CH_2O does have a small primary source from combustion, ambient CH_2O mixing ratios are dominated by secondary production, which occurs via the photochemical oxidation of hydrocarbons. The reaction between hydrocarbons (represented as RH) and the hydroxyl radical (OH) forms organic peroxy radicals (RO_2), shown in reaction I. In the presence of oxides of nitrogen (NO_x = NO + NO_2), RO_2 radicals are converted to alkoxy radicals (RO, reaction II). In the absence of NO_x, conversion of RO_2 to RO occurs via alternate pathways (reaction III).

$$RH + OH + O_2 \xrightarrow{k_1} RO_2 \quad \text{(I)}$$

$$RO_2 + NO \xrightarrow{k_2} RO + NO_2 \quad \text{(II)}$$

$$RO_2 + (HO_2, RO_2) \xrightarrow{k_3} RO \quad \text{(III)}$$

RO radicals can then go on to form CH_2O and an additional radical (in this case, the peroxide radical is shown in reaction IV):

$$RO + O_2 \xrightarrow{k_4} CH_2O + HO_2 \quad \text{(IV)}$$

Because k_4 is typically about three orders of magnitude faster than k_2 and k_3, chemical formation of CH_2O can effectively be written as:

$$RO_2 + (NO, HO_2, RO_2, isomerization) \xrightarrow{k_{eff}} \alpha CH_2O \quad \text{(V)}$$

In reaction V, α represents a branching ratio – that is, the number of CH_2O molecules produced per RO_2 radical. The branching ratio α is different for each VOC and depends on the conversion pathway (i.e. by reaction with NO, HO_2, etc.), and is often used to represent the bulk branching ratio for airmasses. Loss of CH_2O occurs by photolysis and by reaction with OH:

In addition, NO_x also facilitates cycling between HO_2 and OH, effectively speeding up the processes described above:

$$HO_2 + NO \xrightarrow{k_6} OH + NO_2 \quad \text{(VIII)}$$

The formation of NO_2 is also a crucial step in the catalytic formation of O_3. O_3 has a relatively long lifetime (around one week in the troposphere) and tends to accumulate

Wolfe et al. (2016)found that, over a NO_x range of 0.1-2 ppbv, the bulk branching ratio, α, ranged from 0.43 in low NO_x environments to 0.62 in high NO_x environments. They also found that over this range of NO_x conditions, CH_2O formation increased by a factor of 3, and RO_2 formation increased by a factor of 2. Therefore, they concluded that the effect that NO_x imparts on the conversion of HO_2 to OH (i.e. reaction VIII) was the primary driver of the observed relationship between CH_2O and NO_x, and the dependence of α on NO_x plays a lesser (but still important) role.

Because reaction I happens at a much slower rate (orders of magnitude) than reactions II – V, the chemical formation of CH_2O is limited by VOC reactivity (VOCR) and the availability of OH. VOCR is calculated as the sum of the product of the rate constant for the reaction between each VOC with OH ($k_{i,OH}$) multiplied by its concentration, $[VOC]_i$:

$$VOCR = \sum_i k_{i,OH}[VOC]_i \quad \text{(IX)}$$

Valin et al. (2015) found that, because the lifetime of CH_2O is so short at midday (~ 1-3 hours), ambient CH_2O concentrations reflect the rate of hydrocarbon oxidation in an airmass. That is, ambient CH_2O concentrations are in rapid equilibrium with the local oxidative

environment, and CH_2O concentrations will rapidly rise and fall with changes in the rate of hydrocarbon oxidation (i.e. the rate of reaction I – this could be due to changes in the hydrocarbon composition or changes in the oxidative capacity, or both).

Because NO_2 formation (and ensuing O_3 formation) proceeds via the same initial reaction as CH_2O formation, there is a natural link between CH_2O and O_3. That is, environments with high rates of O_3 formation will tend to have relatively high ambient concentrations of CH_2O, and vice versa. However, because of the large disparity in the lifetimes of these two species, they behave differently in the atmosphere and have different time dependencies. In effect, changes in ambient CH_2O are prompted by changes in the local oxidative environment, while O_3 tends to accumulate due to photochemical production throughout the day.

3 Methods

3.1 Overview

DISCOVER-AQ was a four-part field campaign with deployments in Maryland (June-July 2011), California's San Joaquin valley (January-February 2013), Houston, Texas (September 2013), and the northern Front Range area of Colorado (July-August 2014). These four locations were chosen because each is in violation of National Ambient Air Quality Standards (NAAQS) for either O_3 or particulate matter, $PM_{2.5}$ (with respect to O_3, Houston and Colorado are listed as "marginal", northern Maryland is listed as "moderate", and much of California's San Joaquin Valley is listed as "extreme" at the time of this publication). During each deployment, surface measurements were augmented at 6-8 ground sites in the local air quality monitoring network administered by state and local environmental agencies. These sites were situated in the heart of urban areas, in nearby suburban areas, and outlying rural areas. Coordinated aircraft sampling was undertaken to establish vertical distributions of gases above each ground site. Colorado had the highest number of research flights flown (16), and Houston the least (9) with California (10) and Maryland (14) falling in between. The NASA P-3B flew spirals over each ground site, with altitudes ranging from ~300 – 4,000 meters above ground level, and occasional missed approaches where the P-3B descended to ~30 meters. On a typical research flight the P-3B would fly three sorties, spiraling over each site three times (typically in the morning, around midday, and in the afternoon). Typical flight tracks and spiral locations for each phase of DISCOVER-AQ are shown in Figure 1. Because the primary objective of DISCOVER-AQ was to relate column measurements to surface conditions on air quality relevant days, flights were typically conducted under clear-sky or partly cloudy conditions. Airborne measurements of species that are related to satellite-based platforms and are directly related to this work include O_3, NO_2, and CH_2O. Measurements of other species (water vapor, carbon monoxide (CO), methane (CH_4), carbon dioxide (CO_2), nitric oxide (NO), and volatile organic compounds (VOCs)) were used to further aid in analysis. All P-3B measurements used in this work are summarized in Table 1. The only surface-based measurements used in this work were O_3 mixing ratios, which were typically provided by local air quality management districts (i.e. the Maryland Department of the Environment (MDE), the Texas Commission on Environmental Quality (TCEQ), the Colorado Department of Public Health and Environment (CDPHE), the California Air Resources Board (CARB) and the San Joaquin Valley Air Pollution Control District (SJVAPCD)), and were supplemented with measurements from collaborators when needed.

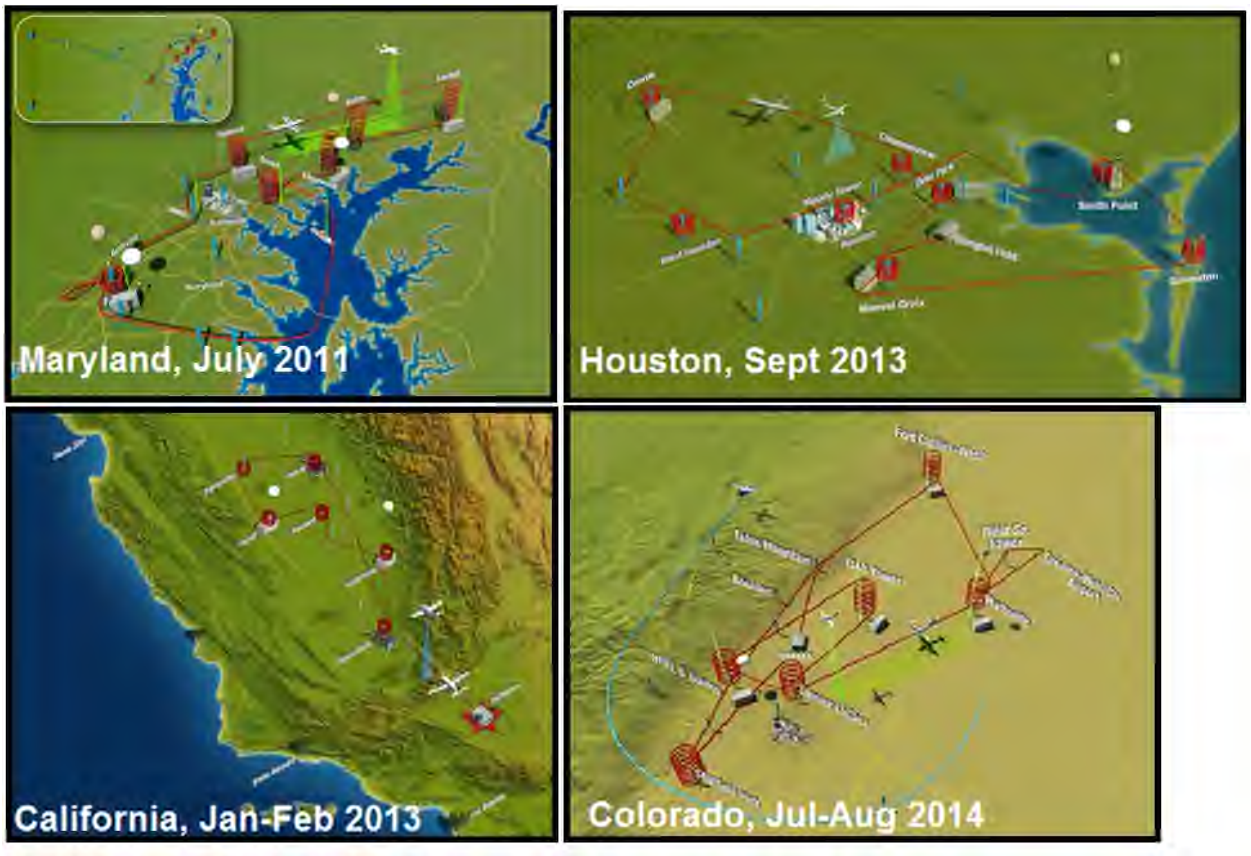


Figure 1. Maps of the study regions of all four DISCOVER-AQ deployments.

Table 1. Summary of P-3B measurements used in this work. All species were measured at a frequency of at least 1 Hz.

Species Measured	**Instrument**	**Reference(s)**
O_3, NO_x	$NO_{xy}O_3$ chemiluminescence	[*Weinheimer et al.*, 1994]
CH_2O	Difference Frequency Absorbing Spectrometer (DFGAS)	[*Weibring et al.*, 2010; *Richter et al.*, 2015]
CO, CH_4	Differential Absorption CO Measurement (DACOM)	[*Sachse et al.*, 1991]
H_2O	Diode Laser Hygrometer (DLH)	[*Diskin et al.*, 2002]
CO_2	Atmospheric Vertical Observations of CO_2 in Earth's Troposphere (AVOCET)	[*Anderson et al.*, 1996; *Vay et al.*, 1999]
VOCs	Proton Transfer Mass Spectrometer (PTrMS)	[*Muller et al.*, 2014]

3.2 Data Analysis – Calculation of Column Densities

Column densities of O_3 and CH_2O were calculated by integrating measurements collected on the P-3B over the altitude range of each spiral. Because the maximum altitude reached during each spiral varied, all spirals were integrated over a standard altitude range (0.3 to 3.2 km above ground level). For spirals where the maximum altitude terminated below the standard range, the remaining portion was calculated by extrapolating the measurement from the top of the spiral up to 3.2 km. Excluding the Beltsville spirals (which had an altitude restriction of 2 km, and therefore are not used in this work), 22% of the remaining DISCOVER-AQ spirals required extrapolation up to the standard range, and the average range of extrapolation was 30 meters (or, about 1% of the total spiral range).

All spirals were extrapolated downwards using mixing ratios measured at the bottom of each spiral (usually ~300 meters AGL) and assuming a constant mixing ratio down to the surface. Flynn et al. (2014) referred to this as the "column_air" approximation, and found that it better-represented the true O_3 column than other methods. On average, the extrapolated amount of CH_2O and O_3 typically made up 13 and 10% of their respective total column densities. On occasion there were missed approaches where the P-3B descended to ~30 meters above ground level. Fried (private communication) compared CH_2O column densities extrapolated from 300 m above the surface to actual results using data from the missed approaches and found that column densities typically only differed by ± 3%.

An additional calculation was performed to calculate the column density of both O_3 and CH_2O within the lowest portion of the column that is chemically perturbed by surface emissions. We refer to this layer as the chemical PBL. Since CH_2O is a product of secondary oxidation of VOC emissions, its vertical gradient does not always follow that might be expected based on the meteorological definition of the PBL (e.g. based on potential temperature gradients). Instead CH_2O often continues to be produced in residual layers above the meteorological PBL, especially in the morning. Because the lifetime of CH_2O is so short (~1-3 hours), profiles exhibit a clear transition between surface influenced values (due to emissions and chemistry) and much lower mixing ratios in the free troposphere. This chemical gradient is the basis for establishing the chemical PBL heights used in this work. Chemical PBL heights were manually determined for each spiral using vertical profiles of CH_2O from each spiral. With these chemical PBL heights established, column densities were calculated using the column_air approach and the chemical PBL height for each spiral as the maximum altitude. Using this approach, the typical extrapolated amount of CH_2O and O_3 (that is, values extrapolated from the surface up to ~ 0.3 km) made up 14 and 17% of their respective chemical PBL column densities. Throughout the rest of this work, these chemical PBL column densities will be referred to as X_{PBL}, where X is either CH_2O or O_3.

3.3 The LaRC Photochemical Box Model

The NASA Langley Research Center (LaRC) time-dependent photochemical box model was used to simulate chemical processes during DISCOVER-AQ. The model is constrained by inputs of trace gas precursors, then calculates the diurnal steady-state profiles of radicals and other computed species (such as CH_2O) for each set of measurements [*Crawford et al.*, 1999; *Olson et al.*, 2001, 2006]. Isoprene chemistry has been updated and is based on the MIM2 scheme [*Taraborrelli et al.*, 2008], with isoprene nitrate chemistry based on Paulot et al., (2009).In effect, the "model time" runs forward until all radical species are in diurnal equilibrium. The

appropriateness of the equilibrium assumption can be a problem when short-lived species dominate the model photochemistry, e.g., highly-reactive VOCs such as biogenic isoprene or alkenes from industrial sources [*Fried et al.*, 2011]. The intricacies of the relationship between model assumptions in proximity to such sources are expanded upon in Section 5 of this work. Additional model uncertainties arise from uncertainties in measured constraints and uncertainties in kinetic and photolytic rates.

When available, model inputs of trace species were calculated using 10-s moving averages of data from the P-3B 1-s data merges. Of the non-methane hydrocarbons (NMHCs) used to constrain the model, ethane, propane, C_{4+} alkanes, ethene, C_{3+} alkenes, ethyne, and higher aromatics were not measured onboard the P-3B. To estimate these missing inputs, data from the Studies of Emissions, Atmospheric Composition, Clouds, and Climate Coupling by Regional Surveys (SEAC^{4}RS) and the Front Range Air Pollution and Photochemistry Experiment (FRAPPE) field campaigns were used. The SEAC^{4}RS instrument payload flew aboard the NASA DC-8 aircraft, and was based out of Houston, Texas from Aug-Sept 2013, and the FRAPPE instrument payload flew aboard the NCAR/NSF C-130 and was based out of Broomfield, Colorado from Jul-Aug 2014. Both field experiments were coincident in space and time with DISCOVER-AQ deployments. During these two campaigns, UC Irvine's whole air sampler (WAS) was used to measure a suite of more than 75 VOCs, including many of the missing NMHC inputs [*Colman et al.*, 2001; *Simpson et al.*, 2010; *Schroeder et al.*, 2014]. To estimate the missing NMHCs in the DISCOVER-AQ datasets, relationships between NMHCs measured by the WAS instrument during SEAC^{4}RS and FRAPPE were used. For example, C_{3+} alkenes were not measured on the P-3B during DISCOVER-AQ, but were measured by WAS during SEAC^{4}RS and FRAPPE. From the WAS data, it was found that the lumped C_{3+} alkene mixing ratio varied linearly with propene mixing ratios. Because propene was measured on the P-3B, this derived relationship was then used to estimate lumped C_{3+} alkenes in the DISCOVER-AQ datasets. Similar relationships were used to estimate mixing ratios of ethane, propane, lumped C_{4+} alkanes, ethene, ethyne, and higher aromatics as well. These relationships are provided in the supporting information accompanying this work. Additionally, total NMHC reactivity (calculated from WAS data) was found to vary linearly with the sum of the mixing ratios of benzene, toluene, propene, and isoprene. This relationship was used to estimate total NMHC reactivity in the P-3B data, and when combined with CH_4 measurements made onboard the P-3B, was used to estimate total hydrocarbon reactivity. The WAS data was also used to estimate the fraction of VOCR that would be required to be estimated in each region. In Maryland, total VOCR was found to be dominated by contributions from CH_4, CO, and biogenic VOCs such as isoprene, and the portion of total VOCR that was estimated using relationships derived from the WAS data was typically less than 20%. In Houston, where anthropogenic emissions were much stronger, this portion was typically less than 35%.

In addition to calculating instantaneous concentrations of radicals and CH_2O, the box model was used to calculate instantaneous rates of formation, destruction, and production for O_3 and CH_2O. Here, "production" is used to refer to the net chemical tendency – that is, difference between the instantaneous rates of formation (i.e. F(X)) and destruction (i.e. D(X) and, as in equation X, where X represents either O_3 or CH_2O:

$$P(X) = F(X) - D(X) \quad \text{(X)}$$

4 Results

4.1 Discover-AQ observations: The Relationship between O_3 and CH_2O

A plot of P-3B column-integrated O_3 versus CH_2O from all four DISCOVER-AQ deployments is shown in Figure 2 in terms of Dobson units (DU), where 1 DU is equal to a column density of 2.69 x 10^{16} molecules/cm^2. A total least squares fit (i.e. orthogonal regression) was applied to these data, and the corresponding coefficient of determination (r^2) is given. While the statistical correlation is fairly modest, this plot shows that high CH_2O column densities are generally associated with high O_3 column densities, and vice-versa. The low r^2 is somewhat driven by the limited dynamic range of the data from the California and Colorado deployments. When only considering Maryland and Houston, the r^2 improves considerably to 0.68.

The anticipated sensitivity for CH_2O detection from TEMPO is ~ 0.35 DU. The anticipated sensitivity for CH_2O detection from TEMPO is ~0.35 DU for a single retrieval (which takes 3 hours). Because most of the data from Colorado and California fall below this cutoff, data from these two locations are not considered in further analysis. It should be noted, however, that resolving information below this precision could be accomplished by averaging measurements either spatially or temporally. Exploring this option to look for O_3-CH_2O relationships over coarser spatial and temporal scales would be interesting, but would requires a larger data set than is available from this study.

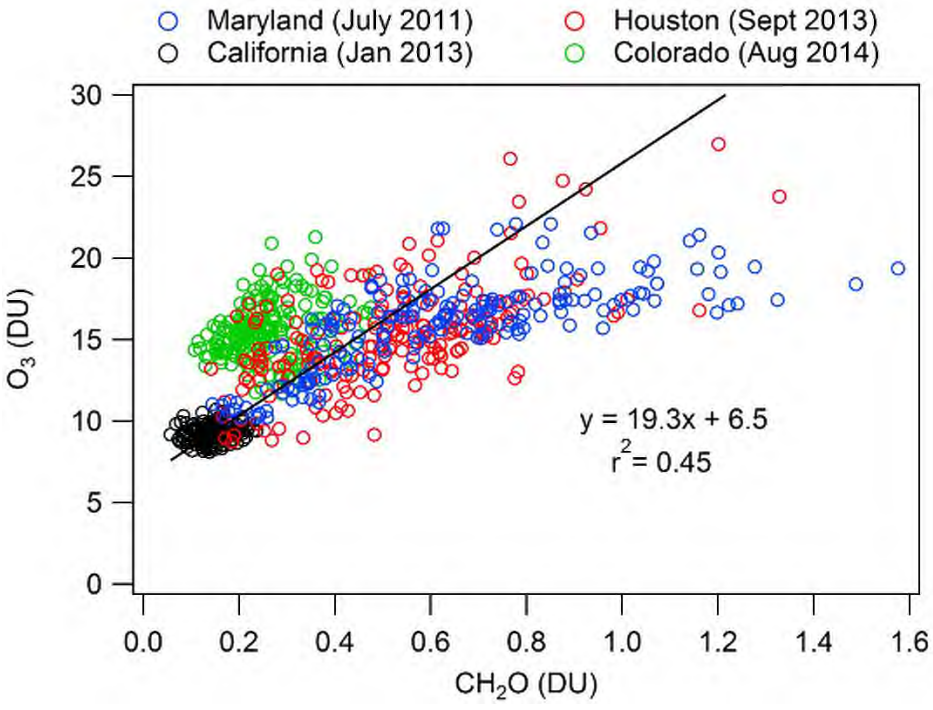


Figure 2. P-3B column-integrated O_3 and CH_2O for each spiral over all sites from all four DISCOVER-AQ deployments.

Figure 3 shows the O3-CH2O relationships for Maryland and Houston when whole column CH2O is compared to PBL column O_3, where both plots show a total least squares fit (i.e. orthogonal regression) rather than an ordinary least squares fit. This relationship (i.e. PBL O_3 vs whole-column CH_2O) was chosen because it is most relevant for air quality monitoring from space – that is, understanding whether total column CH_2O from satellites could be a reliable indicator of near surface O_3 gradients. Fortunately, differences in whole-column and chemical PBL column for CH_2O are small (on average, this subtraction is 0.11 DU, with a maximum of 0.31 DU and a minimum of 0.01 DU) since its lifetime is short and its column abundance and variability is dominated by the chemically perturbed PBL. Considering data from Maryland and Houston in aggregate, the median value for the ratio of $CH_2O_{PBL}/CH_2O_{total}$ was 0.92, while $O_{3,PBL}/O_{3,total}$ was 0.71. Even when considering the remaining column above the maximum altitude sampled by the P-3B, the incremental amount of CH_2O is expected to be small and have less variability than near the surface. In the SEAC4RS dataset, which extends up to 12.5 km in altitude, the median value of CH_2O above 3.2 km was found to be 0.25 ppbv (25^{th} and 75^{th} percentile values of 0.18 and 0.37 ppbv, respectively). Assuming a constant 0.25 ppbv of CH_2O from 3-12 km, we estimate an additional 0.1 DU of CH_2O that would not be accounted for by our spirals.

In the following sections, temporal and spatial trends in the DISCOVER-AQ data are explored (i.e. Sections 4.2 – 4.5), and the LaRC box model is used to show that the relative contribution of biogenic emissions to CH_2O formation explains the differences in behavior seen in Maryland and Houston (Section 5).

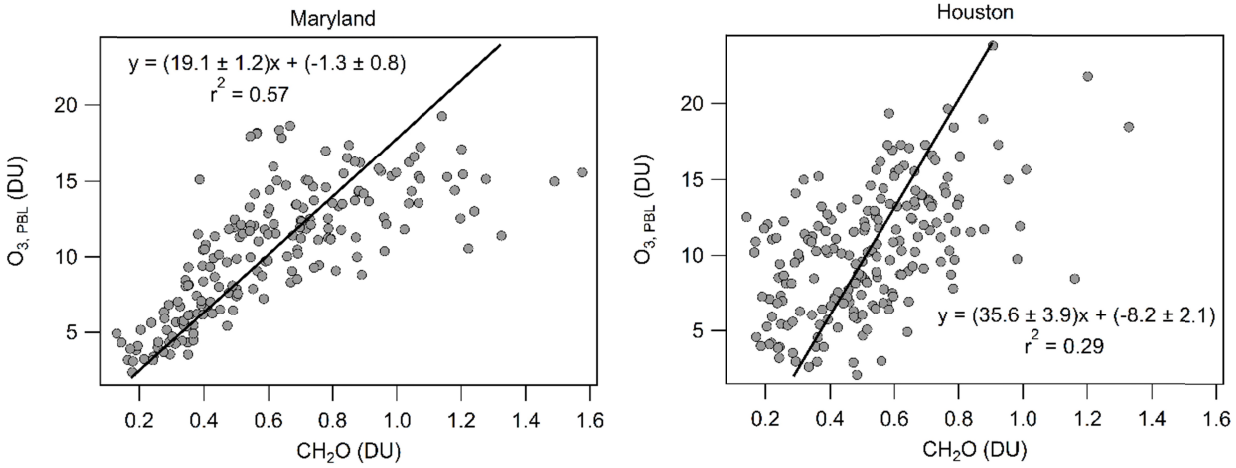


Figure 3. Plots of P-3B integrated $O_{3,PBL}$ and CH_2O column densities for Houston and Maryland. The total least squares fit (i.e. orthogonal regression) and coefficient of correlation are given for each plot.

4.2 Temporal Trends of O_3 and CH_2O in Maryland and Houston

The aggregate statistics for Maryland and Houston show relationships between CH_2O and O_3 that are promising enough to explore in more detail. In order to maximize the effectiveness of future satellite missions such as TEMPO, we must understand the circumstances under which the relationship is most likely to produce useful information. Examining how O_3 and CH_2O vary

spatially and temporally provides insight into which days, and photochemical environments are the best candidates for using column CH_2O measurements to estimate near-surface O_3 tendencies, as well as why some days, and photochemical environments are poor candidates.

Figure 4 shows two individual days (July 28th and 29th) in Maryland. On both days, O_3 reached unhealthy levels in the region (the maximum measured O_3 on each day was ~100 ppbv). On July 28, there was a poor overall correlation ($r^2 = 0.11$) between O_3 and CH_2O, while July 29 had a strong overall correlation ($r^2 = 0.72$). On July 29, CH_2O had a net positive tendency at each spiral site throughout the day. With the exception of Padonia, CH_2O column densities increased from morning to midday to afternoon at each site on July 29th. This can be seen in the bottom right panel, where data points are color-coded by location, and sized by time of day. Here, for each site (except Padonia), CH_2O increased as the day progressed, with the highest values of the day seen during the final spiral at each site. For Padonia, the measured CH_2O column density first decreased between the morning and midday spiral, then increased between the midday and afternoon spiral. This general behavior of monotonically-increasing CH_2O was not seen on July 28. Here, a wide range of CH_2O column densities was observed in the morning, and CH_2O column densities did not monotonically increase throughout the day at any site. This relationship was seen on other days in Maryland as well – that is, days where CH_2O column densities increased throughout the day at each site tended to have stronger overall correlations between O_3 and CH_2O, while days where CH_2O column densities varied little throughout the day at each site tended to have weak overall correlations. Table 2 shows the average CH_2O column density measured in the morning, at midday, and in the afternoon for each research flight, averaged over all locations. Also shown is the average change in CH_2O between morning, midday, and afternoon segments and the r^2 value for the CH_2O-O_3 relationship on each flight day. In general, days where the average change in CH_2O is greater than 0.05 DU (i.e. July 1, 2, 10, 14, 22, 27, and 29) tended to have the strongest correlation between CH_2O and O_3 column densities. In Texas, no days had a visually obvious diurnal increase in CH_2O column density, and generally weak overall correlations between CH_2O and O_3 were observed on each day.

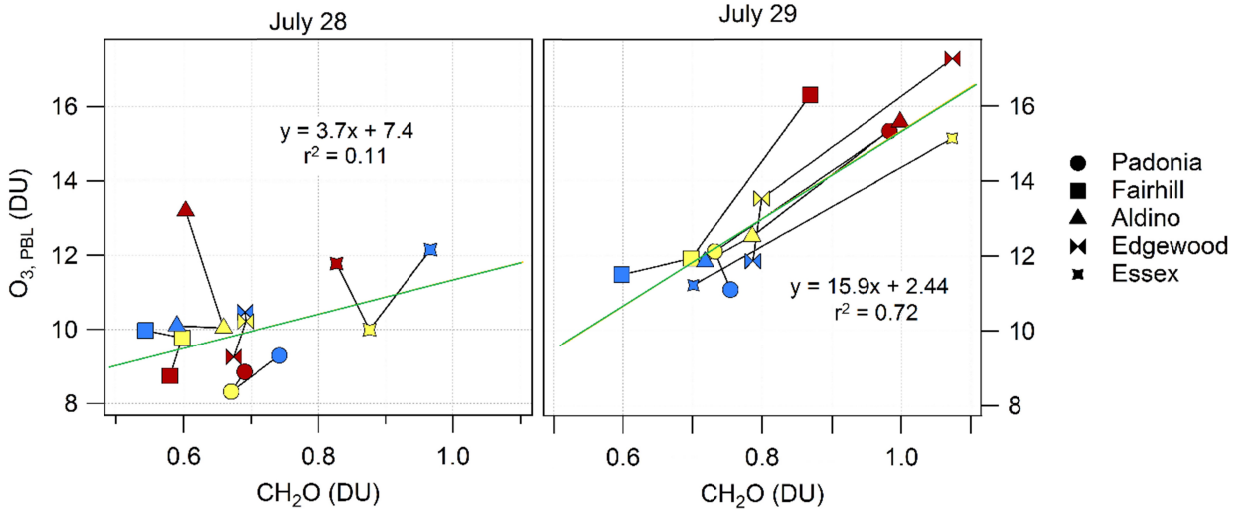


Figure 4. Plots of PBL column O_3 and whole-column CH_2O from July 28 and 29 in Maryland. All points are colored by sortie number (blue = morning spiral, yellow = midday spiral, red =

afternoon spiral), with different shapes corresponding to different spiral locations as noted in the right legend. Least squares regressions are shown as green lines.

Table 2. The average change in CH_2O between segments ((i.e. afternoon-morning)/2) and the r^2 value for the linear fit between O_3 and CH_2O for each flight day in Maryland

Flight #	Date (2011)	Average Change (DU)	r^2
1	July 1	0.105	0.80
2	July 2	0.06	0.78
4	July 10	0.06	0.76
5	July 11	n/a	0.15
6	July 14	0.13	0.83
7	July 16	n/a	0.08
8	July 20	0.05	0.08
9	July 21	0.01	0.42
10	July 22	0.13	0.78
11	July 26	n/a	0.14
12	July 27	0.075	0.70
13	July 28	0.045	0.11
14	July 29	0.095	0.72

From this, it is apparent that estimating near-surface O_3 tendencies from satellite-based CH_2O measurements could be most accurately done on days where CH_2O has a strong diurnal trend. Furthermore, Maryland seems to be a better candidate than Houston for this type of analysis. However, a weak diurnal trend in CH_2O and a poor O_3-CH_2O correlation for an individual day (i.e. the 28th) does not necessarily exclude that day from being included in analysis. When consecutive days are looked at in series, individual days that lack a diurnal scale correlation may contribute important information on a longer-term trend. In Maryland, the P-3B flew on four consecutive days from July 26-29. Over this period, average ambient temperatures increased on each consecutive day. Consequently, biogenic emissions and photochemical activity increased with each consecutive day, and CH_2O followed suit [*Geron et al.*, 2001; *Palmer et al.*, 2006; *Millet et al.*, 2008]. This can be seen in Figure 5, where, although July 28 had a poor correlation, useful information could still be resolved when viewed in the context of the surrounding days.

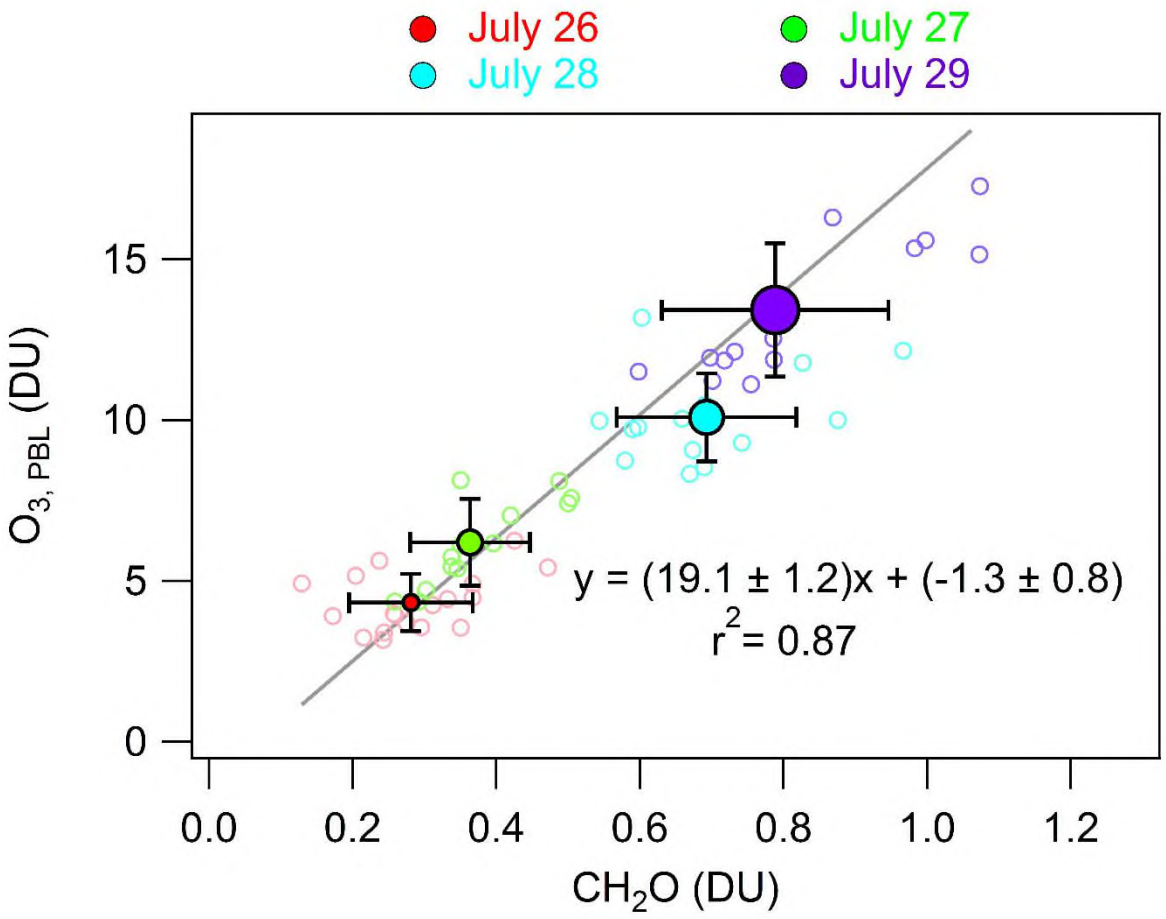


Figure 5. Column CH_2O and chemical PBL-column O_3 over the period July 26-29 in Maryland. Open circles represent each individual spiral, colored by date. Closed circles represent the average values for each day, and variability bars are the 1σ standard deviation. The closed circles are sized by the average column temperature for each day, ranging from 15.9° C on July 26th to 19.9° C on July 29th. (Note: the regression line and r^2 value apply to the individual data points, not the averages)

These observations suggest that column measurements of CH_2O may be useful for estimating near-surface O_3 tendencies (that is, periods of O_3 production) so long as there is sufficient variation in CH_2O – and that the timescale for this variation may be variable. In the next section we explore the O_3-CH_2O relationship when data are filtered based on CH_2O variation. In Section 5 we use modelling and additional observations to explain how diurnal trends in biogenic emissions are the primary driver of diurnal trends in CH_2O, while O_3 production is much less sensitive to changes in biogenic emissions.

4.3 Using ΔCH_2O as a Data Filter

The results in Section 4.2 demonstrate that short-term (i.e., diurnal or multi-day) changes in CH_2O can correspond to associated increases in O_3; however, O_3 exceedances can still occur via local photochemical production without associated changes in CH_2O (for example, if VOCR

stays relatively constant, CH_2O abundances may stay relatively constant while O_3 accumulates), and advection of O_3-rich airmasses could create O_3 exceedances without associated increases in CH_2O. This raises the question of how consistently one could uses changes in CH_2O to indicate periods of O_3 production. To explore this question, data from Maryland and Houston were filtered to identify pairs of consecutive CH_2O observations exhibiting temporal change. The filtering process went as follows: If the CH_2O column density increased by an average of at least 0.025 DU per hour and 5% per hour between two consecutive spirals, both spirals were included. If these criteria were not met, these spirals were not included. For example, in cases where these criteria were met between the 2nd and 3rd spirals but not the 1st and 2nd spirals, only the 2nd and 3rd spirals of that day would be retained. In Maryland, 123 of the original 189 spirals (65%) met these criteria, while 99 of the original 196 spirals (52%) from Houston met these criteria. Plots of CH_2O column densities and O_3 PBL column densities from these filtered datasets are shown in Figure 6.

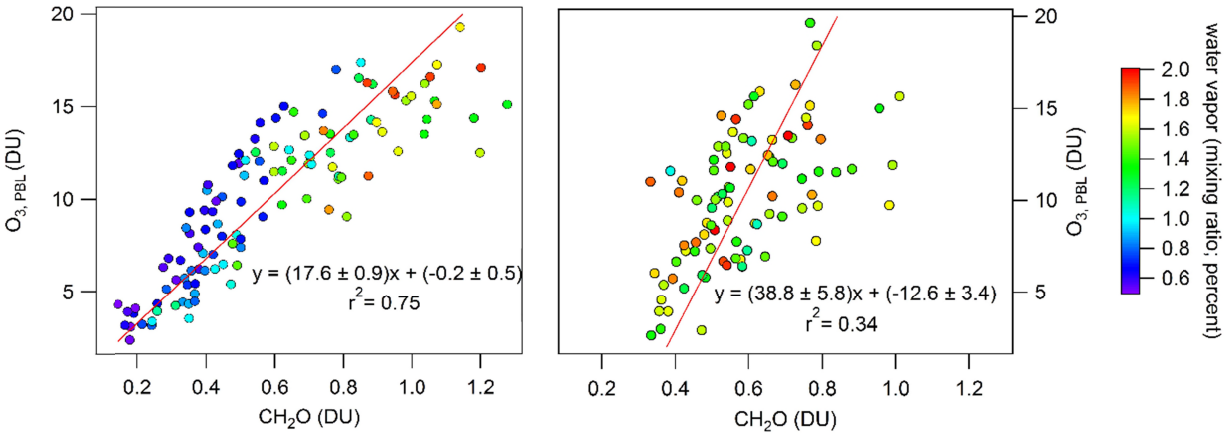


Figure 6. PBL-column O_3 versus column CH_2O in Maryland (left) and Houston (right), filtered to only include columns for which CH_2O changed by at least 0.025 DU and 5% per hour between consecutive spirals. All data are colored by the column average water vapor mixing ratio. Red lines represent the orthogonal least squares fit for each graph.

In Maryland, the statistical correlation greatly improved when this filter was applied (r^2 increased from 0.57 to 0.75), while the correlation in Houston showed only marginal improvement (r^2 changed from 0.30 to 0.34). Applying a filter with stricter requirements (CH_2O had to change by at least 0.05 DU/hr and 10% per hour) had negligible impact on the correlation (r^2 = 0.77 in Maryland, and r^2 = 0.35 in Houston). Thus, application of this filter was effective in Maryland to consistently identify periods of strong O_3 production (that is, to identify periods where O_3 and CH_2O increased concurrently), but it was not sufficient to identify all O_3 episodes. For example, it is possible to have periods where O_3 increases but CH_2O remains flat or decreases (for example, transport of an aged, O_3-rich airmass), which would be removed by this filter. One can also find examples in Houston where early morning CH_2O levels in the immediate vicinity of the ship channel were dominated by primary and pseudo-primary emissions, while O_3 levels at the surface were low due to removal by reaction with NO. For example, CH_2O mixing ratios in excess of 10 ppbv were observed at the bottom of the morning spiral over Deer Park on Sept 25 (~10:30 am), while O_3 showed strong evidence of titration. O_3 mixing ratios at the

surface were less than 20 ppbv until about 9:00, while NO_x was sustained above 60 ppbv. At the time of the P-3B morning spiral at 10:30, O_3 (NO_2) mixing ratios at surface were ~50 (~55) ppbv, were ~55 (~50) ppbv at the bottom of the spiral (300 meters), and were ~65 (~40) ppbv just 500 meters above the surface. In situations such as this, where a large, persistent NO_x source continually titrates O_3 near the surface (especially in the morning hours), changes in surface O_3 are de-coupled from CH_2O and the two are uncorrelated. Using a similar filtering approach to that described above, column CH_2O measurements from future geostationary satellites could be used to identify and monitor the spatial and temporal tendencies of near-surface O_3 production in select locations – but may not be useful in estimating near-surface O_3 mixing ratios or identifying periods of O_3 exceedance.

Data in Figure 6 are colored by the column-average water vapor mixing ratio. This explains some of the deviations from linearity noted in the Maryland data. At high water vapor mixing ratios, the O_3-CH_2O relationship becomes shallower. This is likely because of the primary production of radicals via interactions between water vapor and O_3 (or more accurately, the $O(^1D)$ produced by O_3 photolysis). CH_2O concentrations are often correlated with radical concentrations [*Sillman et al.*, 1995; *Parrish et al.*, 2012], but, because O_3 production in Maryland is strongly NO_x-limited, subsequent increases in radicals (and hence, CH_2O) have a limited effect on O_3 production [*Sillman et al.*, 1990; *Kleinman*, 1994; *Duncan et al.*, 2010]. Additionally, the reaction between $O(^1D)$ and water vapor acts as a sink of O_3, resulting in a lower net $P(O_3)$ at high water vapor loadings, all else being equal.

It is important to note that this filter isolates short-term periods of O_3 production and should not preclude examining CH_2O variability over periods of multiple days as a useful way to show changes in air quality associated with synoptic-scale influences, as in the case study of July 26-29 presented in section 4.2. At present, however, no judgment can be made on the further applicability of the results of that case study because the DISCOVER-AQ data lacks sufficient examples of multi-day fights with increasing (or decreasing) CH_2O. In Houston, the P-3B sampled four consecutive days on one occasion (Sept 11-14), but the daily average CH_2O column density only varied by ~0.2 DU over these four days, and did not vary monotonically. There were two other events (July 20-22 in Maryland and Sept 24-26 in Houston) where the P-3B sampled on three consecutive days, but the daily average CH_2O column density did not vary monotonically over these three days – in both cases, it increased from the first to the second day of the period, then decreased from the second to the third day, indicating that those periods did not cover a lone synoptic-scale air quality event that spanned multiple days. In effect, when the timespan of an individual air quality event spans multiple days (as in July 26-29), data collected throughout the duration of the event may be viewed collectively. However, when an air quality event only spans one day, aggregating data over periods of multiple days is not appropriate. In the next section, we show that persistently strong spatial gradients in Houston mask any temporal trends in the O_3-CH_2O relationship, and in Section 5 we show that differences in the relative contributions of biogenic and anthropogenic VOC emissions to secondary CH_2O formation explains the differences in behavior between Maryland and Houston.

4.4 Spatial Trends in the Column O_3-CH_2O Relationship

In the previous sections, it was established that O_3 PBL columns correlate better with column CH_2O in Maryland than in Texas. Part of this may come from the reduced dynamic range of air quality conditions observed in Houston compared to Maryland. During the Maryland campaign,

O_3 exceedances were recorded on 9 of the 14 flight days. Additionally, two of the remaining flight days were exceedingly clean due to a period of strong transport from Canada. By contrast, conditions were stubbornly consistent during the Houston deployment, with cloudy, often rainy, and moderately polluted conditions (but no O_3 exceedances) persisting until the final flight-days of the deployment on 25-26 September. On 25 September in particular, Houston recorded a one-hour O_3 exceedance in excess of 120 ppbv at the Manvel Croix site, and exceedances were recorded widely from the Houston ship channel to the south, whereas conditions were clean to the north. This strong spatial contrast was created due to a bay breeze recirculation of pollution from the ship channel over areas to the south, and very strong emissions of highly-reactive VOCs, such as propene, ethene, and isoprene, from the ship channel region [*Loughner et al.*, 2014; *Fried et al.*, 2016].

By looking at the data for 25 September spatially rather than temporally, some interesting relationships are revealed. Overall there was a weak correlation between PBL-column O_3 and column CH_2O (i.e. the dashed black line in Figure 7). However, when data from the midday and afternoon circuits were isolated (i.e. the orange and red lines in Figure 7), stronger correlations emerged, and the slope increased from 8.1 at midday to 11.6 in the afternoon. Due to the bay breeze recirculation mentioned above, conditions on this day were fairly stagnant with a light northwesterly breeze in the morning, and a light southeasterly sea breeze in the afternoon. The spatial influence on CH_2O is evident in that the highest column values were restricted to the ship channel area (i.e., Deer Park and Channelview, where CH_2O column densities were greater than 0.8 DU) which was dominated by anthropogenic emissions. Even at Channelview, just to the north of the ship channel, high CH_2O from ship channel emissions was not present until the second circuit. By the third circuit, emissions were seen to reach Moody Tower and Manvel Croix as column values approached 0.8 DU. Only one other site (Smith Point, located southeast of the ship channel) saw CH_2O approaching 0.8 DU, but it was on the second circuit prior to the full build-up of O_3.

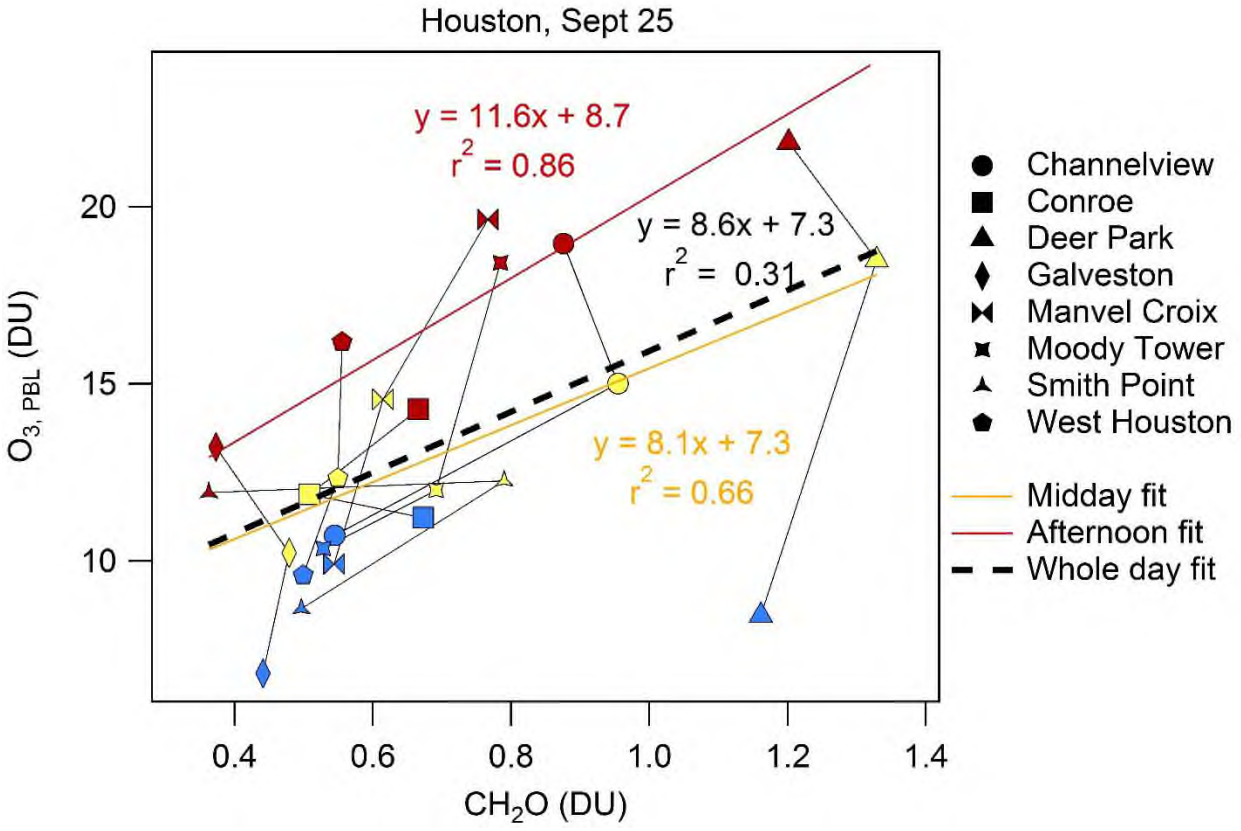


Figure 7. PBL-column O_3 versus column CH_2O in Houston on Sept 25th. All points are colored by sortie number (blue = morning, yellow = midday, and red = afternoon) and with shapes corresponding to different locations as noted in the legend. The red line and equation show the least squares fit for afternoon data only, orange for midday only, and black for all data points.

Trajectory calculations done using the NOAA HYSPLIT dispersion model (http://ready.arl.noaa.gov/HYSPLIT_disp.php) reveal that the ship channel plume reached Smith Point by midday, but then dissipated by the afternoon. Thus, the large increase/decrease pattern at Smith Point on this day was likely due to transport from the ship channel. In the case of Manvel Croix, it took several hours for the ship channel plume to reach the local area, and it arrived just before the midday spiral. For Moody Tower, the plume also took several hours to reach the local area, but it arrived between the midday and afternoon spirals. In both cases, there was a large increase in O_3 upon arrival of the plume, but a modest increase in CH_2O (still larger than the filter cutoff described in the previous section, though). VOCR at the ship channel was dominated by emissions of short-lived VOCs such as propene, ethene, and isoprene [*Murphy and Allen*, 2005; *Vizuete et al.*, 2008; *Parrish et al.*, 2012; *Zhu et al.*, 2014; *Fried et al.*, 2016]. After several hours of transport, these short-lived VOCs would be fairly depleted, and consequently the ambient CH_2O mixing ratio would be reduced as well [*Fried et al.*, 2011]. Conversely, O_3 would have accumulated throughout the whole transport process. Thus, depending on how long the transport time is between the ship channel and a given location, the ratio of $\Delta O_3/\Delta CH_2O$ upon plume arrival will vary in turn. This effect has been previously observed in areas downwind of large VOC emission events in Houston. Parrish et al. (2012) and Wert (2003) both found that, during downwind transects of the ship channel plume, the CH_2O/O_3 ratio peaked in the near field

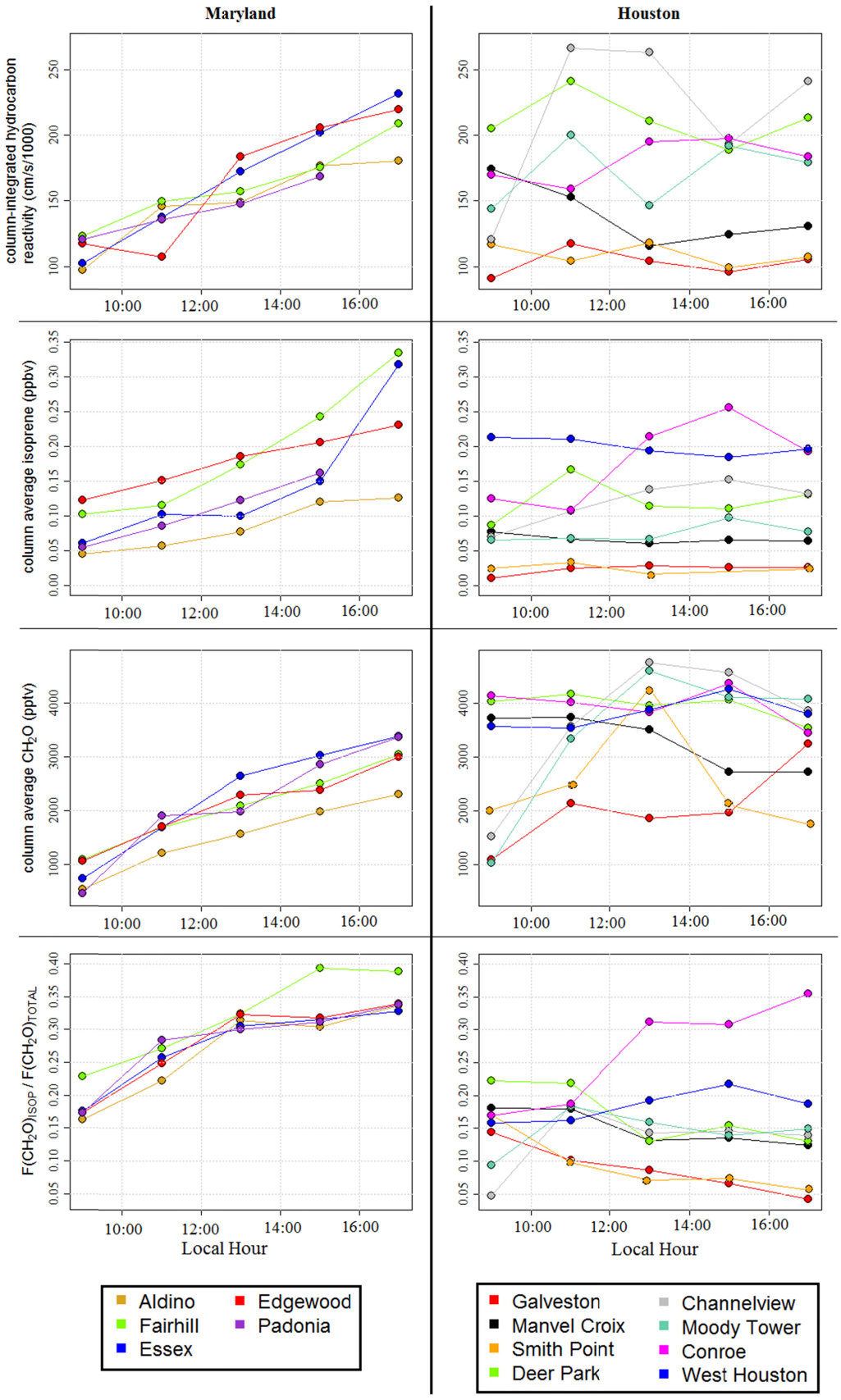
(about 10-20 km downwind, or ~1 hour of transport time), and then decreased as the plume moved further downwind. This is likely why the filter approach shown in Figure 6 failed to improve the correlation in Houston: Spirals conducted in areas that were far downwind of the ship channel plume may still have experienced a large enough change in CH_2O to be above the filter cutoff, but had a disproportionately large increase in O_3, thereby introducing another source of uncertainty in the form of transport time. In Maryland, where biogenic emissions (which are spread out throughout the region) dominate the regional VOCR, this issue does not arise.

5 Discussion

While the observations presented in Section 4 alone may be useful, they do not yet explain the discrepancy between Maryland and Houston - that is, what might be the theoretical basis for why the O_3-CH_2O correlation is much stronger in Maryland than in Houston. In the previous section, it was shown that temporal variability in CH_2O column densities was required for a strong correlation to emerge, and it was hypothesized that differences in biogenic vs anthropogenic contributions to the local hydrocarbon mix was the cause of the discrepancy between Maryland and Houston. In Section 5.1, this idea is further developed and observational evidence is provided to show that a widespread, temporally-variable biogenic source dominates CH_2O production in Maryland, while near-constant anthropogenic sources dominate CH_2O production in Houston. In this section, the LaRC photochemical box model is also used to show that primary CH_2O emissions likely contributed a minimal amount to the CH_2O abundances observed during DISCOVER-AQ. In Section 5.2, the LaRC photochemical box model is employed to provide a theoretical understanding of why O_3 and CH_2O best correlate when CH_2O has strong temporal variability.

5.1. The Impacts of Biogenic and Anthropogenic Hydrocarbon Emissions on CH_2O in Maryland and Houston

While secondary production of CH_2O can be affected by abundances of NO_x and other radical species (i.e. the branching ratio α in equation V), ambient CH_2O mixing ratios are primarily a function of the ambient VOCR [*Parrish et al.*, 2012; *Valin et al.*, 2014; *Wolfe et al.*, 2015]. Hourly averages of column-integrated hydrocarbon reactivities (that is, column-integrated VOCRs; calculated using the WAS correlations shown in the supporting information accompanying this work - It is of note that this approach excludes the contribution from OVOCs, which can make up a significant portion of total OH reactivity.) for spiral sites in Maryland and Houston are shown in the top two panels of Figure 8. Column-integrated reactivities tended to increase throughout the day at all Maryland sites, while Houston sites tended to increase in the morning hours then remain relatively flat the rest of the day.


Maryland
Houston
column-integrated hydrocarbon reactivity (cm/s/1000)
column average isoprene (ppbv)
column average CH_2O (pptv)
$F(CH_2O)_{ISOP} / F(CH_2O)_{TOTAL}$
Local Hour
Aldino
Edgewood
Fairhill
Padonia
Essex
Galveston
Channelview
Manvel Croix
Moody Tower
Smith Point
Conroe
Deer Park
West Houston

623

Figure 8. Typical diurnal trends of column-integrated hydrocarbon reactivity (top panels, divided by 1000), PBL-column-average isoprene (second from top), PBL-column average CH_2O (second from bottom), and the PBL-column-average fractional contribution of isoprene to CH_2O formation (bottom) in Maryland (left) and Houston (right). All data were binned into two-hour time bins, and the median of each bin is shown for each site. Column average isoprene in Conroe has been divided by 3 to keep scaling consistent.

In Houston, the hydrocarbon mix is dominated by anthropogenic emissions with some influence from biogenic sources. Of particular importance are refineries near the ship channel that emit highly reactive NMHCs like ethene, propene, and anthropogenic isoprene [*Ryerson*, 2003; *Murphy and Allen*, 2005; *Vizuete et al.*, 2008; *Zhu et al.*, 2014; *Fried et al.*, 2016]. By contrast, the hydrocarbon mix in Maryland is largely dominated by a strong biogenic source with an appreciable contribution from anthropogenic emissions. This difference in hydrocarbon emissions can explain the difference in the diurnal behavior of column VOCR between the two locations. The second-from-the-top panels of Figure 8 show the diurnal behavior in column-average isoprene mixing ratios for Maryland and Houston. For Maryland, isoprene increased throughout the day while Houston exhibited lower isoprene levels with no consistent diurnal behavior. One could speculate that this is due to the fairly cloudy conditions during many of the Houston flights - which could have suppressed biogenic emissions - or the fact that the Houston deployment took place in September, when isoprene emissions typically begin to decrease [*Geron et al.*, 2001; *Palmer et al.*, 2006; *Potosnak et al.*, 2014]. This is inconsistent, however, with isoprene behavior at the Conroe site, located north of Houston in a more wooded area. It is more likely that a strong source of anthropogenic isoprene was observed over the refineries near the ship channel in Houston. Because these refineries are active throughout the day, isoprene in the region had little diurnal variability.

Globally, isoprene is the largest VOC contributor to secondary CH_2O formation. To estimate the role of isoprene in driving changes in ambient CH_2O mixing ratios for Maryland and Houston, the fractional contribution of isoprene and its oxidation products to the instantaneous rate of CH_2O formation was calculated using the LaRC photochemical box model. This quantity was calculated as the rate of formation of CH_2O from isoprene and its oxidation products [*Palmer et al.*, 2003; *Millet et al.*, 2006; *Marais et al.*, 2012]. The value reported here is the PBL-column-average of $FCH_2O_{ISOP}/FCH_2O_{total}$. Looking at all data collected in Maryland, the median $FCH_2O_{ISOP}/FCH_2O_{total}$ between 250 and 500 meters (i.e. the bottom of each spiral) was 0.75, but decreased to 0.2 at an altitude of 1 km. This is because the vertical distributions of isoprene and its oxidations products are heavily skewed towards the surface – the median isoprene mixing ratio between 250 and 500 meters in Maryland was 0.5 ppbv, but fell to 0.1 ppbv at 1 km. Therefore, the PBL-column-average values of $FCH_2O_{ISOP}/FCH_2O_{total}$ that we report here are lower than the 0.5 – 0.8 reported by studies that focused on near-surface chemistry in Eastern North America [*Lee et al.*, 1998; *Macdonald et al.*, 2001; *Sumner et al.*, 2001], but are consistent with the 0.2 – 0.3 reported by Pfister et al. (2008), who looked at the average contribution of isoprene to CH_2O formation in the vertical column. Diurnal trends in column averages of CH_2O and $FCH_2O_{ISOP}/FCH_2O_{total}$ are shown for all sites in Maryland and Houston in the bottom two rows of panels in Figure 8. In Houston, $FCH_2O_{ISOP}/FCH_2O_{total}$ had less variation than in Maryland, and ambient CH_2O mixing ratios tended to remain relatively flat throughout the day. This, coupled with the weak diurnal trend in isoprene abundances in Houston and the overall lower PBL-column-average $FCH_2O_{ISOP}/FCH_2O_{total}$ (~15% in Houston versus

669 ~28% in Maryland) suggest that biogenic emissions play a proportionately smaller role in
670 driving changes in ambient CH_2O levels in Houston. This is in agreement with previous studies
671 that have found anthropogenic emissions to be the leading contributor to secondary CH_2O in
672 Houston [*Wert*, 2003; *Parrish et al.*, 2012; *Zhu et al.*, 2014]. It is also worth noting that there is
673 an anthropogenic source of isoprene from the refineries in the Houston ship channel, so the true
674 biogenic contribution to CH_2O formation is likely smaller than reported here. In Maryland, a
675 clear relationship is observed: column average CH_2O mixing ratios increase (by about 250 pptv
676 per hour, on average) as column-average isoprene mixing ratios increase (by about 25 pptv per
677 hour), and the fractional contribution of isoprene increases (by about 0.02 per hour). This,
678 coupled with the positive diurnal tendency in isoprene in Maryland suggests that changes in
679 biogenic emissions throughout the day play a large role in driving changes in ambient CH_2O
680 levels. However, it should be noted that even if all five carbon atoms from every isoprene
681 molecule went on to form CH_2O, the expected increase in CH_2O due to changes in isoprene
682 alone would be about 125 pptv per hour (or about 1 – 1.5 ppbv over a day) – about half of the
683 observed increase of 250 pptv per hour (or about 2 - 3 ppbv over a day). This suggests that while
684 biogenics play an important role in driving diurnal trends in CH_2O in Maryland, there is another
685 significant factor that must account for an increase of at least 1 ppbv of CH_2O throughout the
686 day. The identity of this "missing" factor can be explained by the diurnal tendency of CH_2O in
687 the absence of diurnally-varying biogenic emissions, which is discussed in greater detail in
688 Section 5.2.

689 While ambient CH_2O mixing ratios are dominated by secondary production, an
690 understanding of the relative contribution of primary CH_2O emissions is needed in order to fully
691 understand these observations. Because the LaRC photochemical box model only includes
692 secondary production of CH_2O, the degree of agreement between measured and modeled CH_2O
693 concentrations can be used as an indicator of the importance of primary emissions. Figure 9
694 shows a comparison of modeled and measured CH_2O mixing ratios in Houston on the afternoon
695 of September 25th. On this day, the sky was clear and conditions were hot and fairly stagnant.
696 Under these conditions, O_3 production was high and O_3 mixing ratios in excess of 100 ppbv were
697 observed throughout the region, with a maximum of 145 ppbv measured south of downtown.

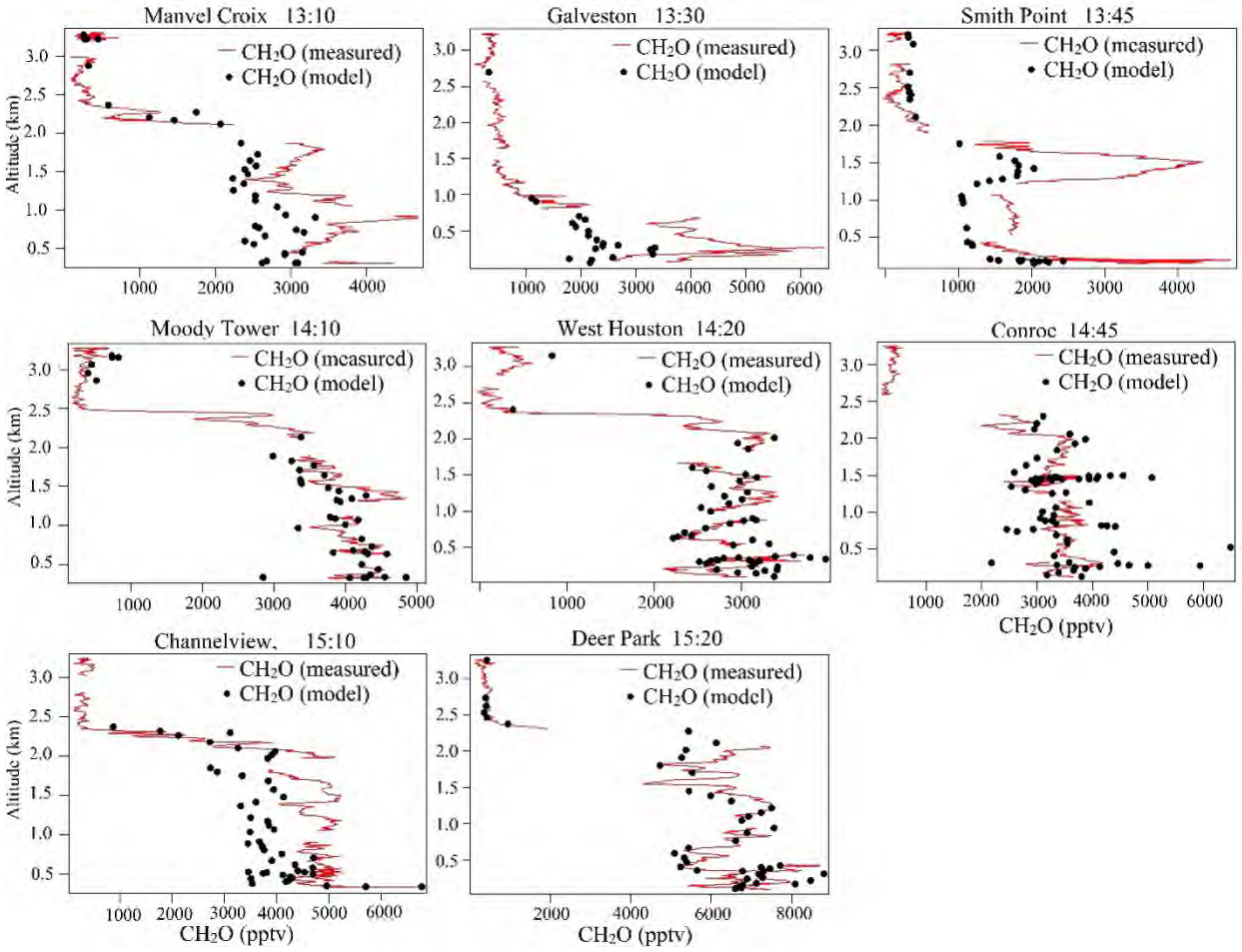


Figure 9. Comparison of model-predicted CH_2O (black dots) and measured CH_2O (red lines) for the third circuit on Sept 25 in Houston. For both traces, a moving 10-s average is shown. The y-axis has the same scaling for each panel, but the x-axis differs for each panel.

In Figure 9, model-predicted CH_2O mixing ratios are generally within 20% of measured CH_2O mixing ratios during each spiral on the third circuit. Two notable exceptions are Galveston and Smith Point, where the model under-predicted CH_2O mixing ratios. The model is also low-biased in Channelview, but still within 20% of measurements. Because the lifetime of a given molecule will determine the length of time it will take to reach equilibrium, the time-distance from an emission source determines which species will be accurately predicted by the model. For example, when near a short-duration emission source (i.e. a "puff"), the model will most-accurately predict species with very short lifetimes and will tend to over-predict species with longer lifetimes. When near a persistent emission source, such as a power plant or petrochemical facility, the model will generally be in good agreement with measurements except when strong advection occurs. This scenario is similar to September 25 in Houston, which was characterized by strong, persistent emissions from the ship channel, low wind speeds, and weak advection. Here, the model performed best in urban areas and when near emission sources, such as Moody Tower, West Houston, Conroe, Channelview, and Deer Park. When far downwind of an emission source, the model may under-predict species with longer lifetimes due to the lack of ability to account for the time-history of an airmass. That is, the model cannot accurately account for previous accumulation of species with longer lifetimes. Although CH_2O is considered a

short-lived molecule, its lifetime is sufficiently long that it is susceptible to the deviations from equilibrium described above. This is why the model is low-biased in downwind sites such as Galveston and Smith Point, and to some extent Manvel Croix, which have chemical compositions that are dominated by transport rather than local emissions. Previous work has found that model predictions of CH_2O in well-aged airmasses may be under-predicted by a factor of 1.6 [*Fried et al.*, 2011].

On other hot, clear-sky days in Houston and Maryland, similar agreement between model and measurement was observed (not shown, though generally within 30%). This indicates that on days where photochemical production of O_3 is high, primary sources of CH_2O likely contribute a minimal amount to observed ambient CH_2O mixing ratios. As noted above, 0-D box models may over-predict or under-predict CH_2O mixing ratios under certain situations. Thus, while our results suggest a minimal importance of primary CH_2O emissions on hot, clear-sky days, they do not completely exclude the possibility that primary CH_2O emissions are important on these days. However, many literature studies are in agreement: In very large urban areas such as New York City [*Lin et al.*, 2012b] and Mexico City [*Garcia et al.*, 2006], photochemical production accounts for ~70% of observed ambient CH_2O mixing ratios at midday. In smaller urban areas, such as Columbus, Ohio, secondary CH_2O production account for 80% of observed ambient CH_2O [*Mukund et al.*, 1996], and it is estimated that, on average, secondary production of CH_2O accounts for more than 95% of observed CH_2O in the eastern half of the U.S. [*Li et al.*, 1994]. Further complicating this issue, separation of true primary emissions from pseudo-primary emissions (that is, photochemically-produced CH_2O that is created in the immediate vicinity of an emission source) is very difficult. Many studies use correlations between CH_2O and CO (a tracer of combustion) to determine the portion of CH_2O attributed to primary emissions. However, highly-reactive VOCs are also co-emitted during the combustion process, and in the presence of these highly-reactive VOCs, large amounts of CH_2O can be formed very quickly [*Fried et al.*, 2016]. Measurements collected even a short distance downwind of a suspected primary source of CH_2O may be likely to over-predict the amount of primary CH_2O due to unaccounted for pseudo-primary CH_2O produced between the source and the measurement site. Airborne studies conducted in Houston have explored this idea, and also concluded that it is very difficult to separate primary CH_2O emissions from pseudo-primary emissions. These studies also concluded that true primary emissions of CH_2O likely contribute a negligible amount to ambient CH_2O mixing ratios [*Wert*, 2003; *Parrish et al.*, 2012; *Fried et al.*, 2016]. Thus, the remainder of this paper will operate under the assumption that primary emissions are not a major contributor to the observed variability in CH_2O column densities.

5.2 A Theoretical Basis for Understanding the O_3-CH_2O Relationship

Though O_3 and CH_2O are produced by similar photochemical reactions, it is unlikely that co-production is the reason for this correlation. As described in Sections 1 and 2, O_3 is expected to accumulate throughout the day while ambient CH_2O mixing ratios are generally expected to reflect the local hydrocarbon oxidation environment. To explore this hypothesis, the LaRC photochemical box model was employed to investigate diurnal trends for O_3 and CH_2O. Two separate simulations were run – one using average trace gas inputs from below 1.5 km in Maryland, and the other using average trace gas inputs from below 1.5 km in Houston. In both cases, the model was initialized at noon local time, and these trace gas inputs were held constant until the end of the following day, while O_3, CH_2O, and all radical species were allowed to vary. The results of these simulations (shown in Figure 9) again suggest different chemical behaviors

764 in Maryland and Houston. In both simulations, O_3 increased throughout the day (by about 20
765 ppbv in Maryland, and about 55 ppbv in Houston), while CH_2O had very different tendencies
766 between the two locations. In Maryland, CH_2O reached a daily minimum of 3.9 ppbv near
767 sunrise, then increased monotonically throughout the day (with a higher rate of increase in the
768 morning than the afternoon) until reaching a maximum value of 4.9 ppbv at sunset. This 1 ppbv
769 increase in CH_2O throughout the day due to chemical production under constant VOC conditions
770 may explain the "missing" CH_2O noted in Section 5.1. As noted in Section 5.1, isoprene and the
771 total VOCR tended to increase throughout the day in Maryland, so the 1 ppbv of CH_2O produced
772 in this simulation where VOCs were held constant likely represents a low estimate of the net
773 accumulation of CH_2O from photochemical processing. In Houston, CH_2O reached a daily
774 minimum of 4.4 ppbv at sunrise, had a high rate of increase in the early morning, reached a daily
775 maximum of 5.5 ppbv in the late morning, then decreased to a local minimum of 5.1 ppbv at
776 around 16:00 local time. These differences in behavior can be attributed to differences in the
777 average chemical composition between Maryland and Houston. Below 1.5 km, the aggregate
778 mixing ratios of most VOCs were higher in Houston than in Maryland (with the notable
779 exception of isoprene, which was 45% higher in Maryland), and the aggregate NO_x mixing ratio
780 in Houston was 136% higher than in Maryland (2.81 ppbv in Houston versus 1.19 ppbv in
781 Maryland). Because of the large role that NO_x plays in cycling between HO_2 and OH (i.e.
782 reaction VIII), simulated midday mixing ratios of OH in Houston were a factor of 2.5 higher than
783 in Maryland. As a result, the midday lifetime of CH_2O due to reaction with OH in Houston was a
784 factor of 2.5 lower than in Maryland (2.1 hours in Houston versus 5.3 hours in Maryland). For
785 reference, the midday lifetime of CH_2O due to loss by photolysis was about 3.7 hours in both
786 locations. Therefore, the overall lifetime of CH_2O (that is, including loss by photolysis and
787 reaction with OH) at midday in Houston was lower than in Maryland by a factor of 1.7. As a
788 result of this short lifetime, CH_2O had very little tendency to accumulate in Houston, and
789 responded much more quickly to changes in the rate of hydrocarbon oxidation than in Maryland.
790 Because hydrocarbons were effectively held constant in these simulations, variability in the rate
791 of hydrocarbon oxidation could only arise from changes in OH – which has a diurnal profile that
792 is nearly identical to that of J_{NO2} (i.e. the green trace in Figure 10). When coupled with the fact
793 that reaction with OH was the dominant removal pathway for CH_2O in Houston, an
794 understanding of the unique diurnal trend in CH_2O observed in Houston emerges. In the early
795 morning, CH_2O increased in response to initial increases in OH – but, by the late morning, the
796 CH_2O lifetime became short enough such that accumulation was retarded, and the CH_2O mixing
797 ratio began decreasing by midday as a result of increased destruction.

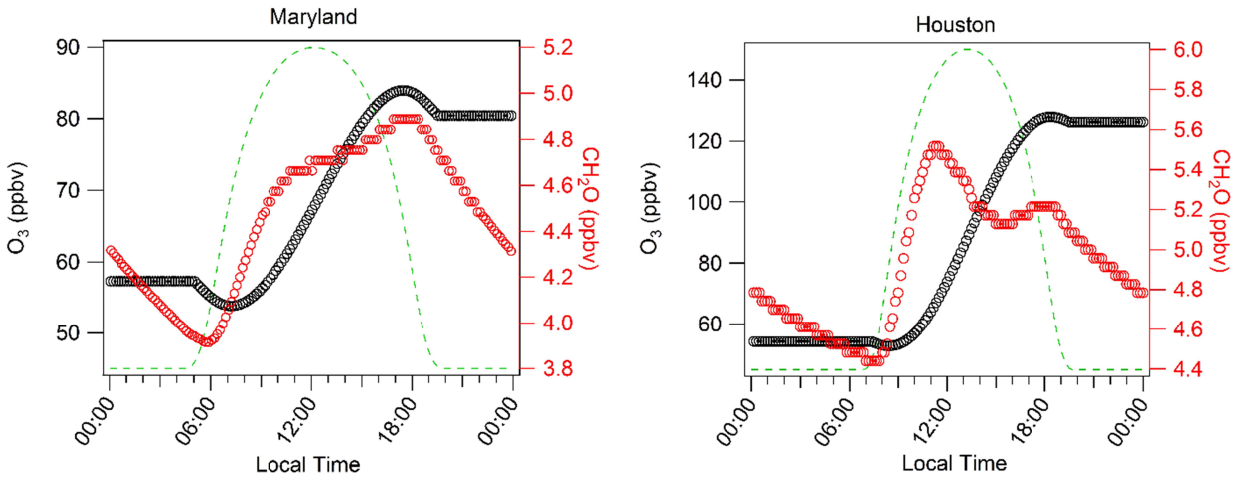


Figure 9. Model results for average chemical conditions in Maryland (left) and Houston (right). Average VOC and trace gas mixing ratios from each location were used as inputs and held constant throughout the day. O_3 (black trace), CH_2O (red trace), and other radical species (not shown) were allowed to vary throughout the day. The green trace is unitless and is a qualitative representation of the actinic flux.

To explore how changes in airmass composition affect equilibrium CH_2O mixing ratios, a "base case" diurnal profile of CH_2O and O_3 in relatively clean conditions was created where trace gas inputs were held constant throughout the day. Then, three additional simulations were run such that at 12:00, the airmass composition was abruptly changed. The results of these simulations are shown in Figure S8 in the supplemental information accompanying this work, where the base case is shown as a black trace, and the three simulations where composition was changed are represented as traces A, B, and C. The NO_x, alkane, and alkene input mixing ratios for the base case and simulations A, B, and C are given in Table S1. In all cases, when the airmass composition was changed, CH_2O mixing ratios rapidly re-adjusted to new equilibrium levels within about three hours. The amount of CH_2O present once equilibrium was re-established, however, depended on the new airmass composition. Note that the diurnal profile for CH_2O in this simulation appears slightly different from those in Figure 9 – this is because the lower amount of CH_2O present in the base case simulation yields lower rates of CH_2O destruction, which is first order with respect to CH_2O.

These results imply that changes in ambient O_3 concentrations can be decoupled from changes in ambient CH_2O concentrations. In a regime where primary CH_2O emissions are negligible and emissions of chemical precursors remain relatively constant throughout the day, CH_2O is expected to reach equilibrium in the morning, and have small fluctuations throughout the rest of the day – in the constant-VOC simulations presented in Figure 10, the daily maximum CH_2O mixing ratio was only about ~20% higher than the daily minimum. At the same time, if O_3 production is positive, we expect O_3 to increase monotonically throughout the day by virtue of its long lifetime. On the other hand, in an O_3-producing regime where precursor emissions increase throughout the day, we expect to see O_3 and CH_2O co-vary with a much larger dynamic range in CH_2O than was observed in Figure 10. To test this hypothesis, another simulation was run using the base case constraints described above. In this simulation, however, isoprene was set to increase at a rate of 600 pptv per hour throughout the daytime hours. The results of this simulation are shown in Figure 10. Here, CH_2O increased fairly linearly at a rate of ~1 ppbv per

hour in response to increasing VOCR from isoprene. Calculated values of $P(O_3)$ increased until reaching a maximum value of ~6 ppbv/hour before midday, then began to decrease again at 16:00. In this environment, where VOCR increased throughout the day, $P(O_3)$ was limited by NO_x availability and thus was not affected by changes in isoprene to the same degree that CH_2O was.

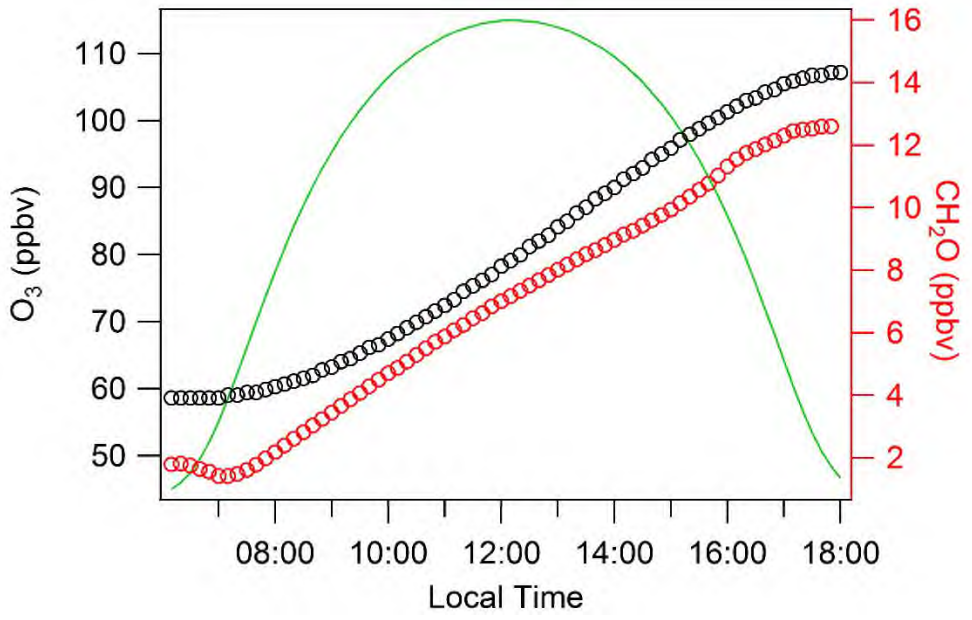


Figure 10. Box model simulation of O_3 (black) and CH_2O (red), using the inputs described in the base case scenario above. During daytime hours, isoprene was set to increase at a rate of 600 pptv per hour. The green trace is unitless and is a qualitative representation of actinic flux.

The results of these simulations, along with the observations shown in Section 5.1, provide support to the hypotheses suggested in Section 4: O_3 and CH_2O column densities have a strong correlation only when temporal variability in CH_2O is introduced, and better correlations were observed in the biogenically-dominated environment of Maryland because of the temporal variability that biogenic emissions imprinted on ambient CH_2O values. However, these simulations do not provide a mathematical framework for understanding the results presented in Section 4. For example, these simulations do not explain why a slope of ~13 DU O_3/DU CH_2O is observed in Maryland, nor do they explain the observed increase in slope between midday and afternoon when data from Houston were viewed spatially. What follows is a mathematical description of these results, derived from first principles.

Because of the short lifetime of CH_2O, mixing ratios are essentially proportional to the instantaneous rate of formation of CH_2O ($F(CH_2O)$). Because $F(CH_2O)$ is related to the local oxidative environment (i.e. a function of actinic flux, VOCR, OH, and NO_x), $F(CH_2O)$ is therefore proportional to the instantaneous rate of formation of O_3 ($F(O_3)$). Because O_3 has a relatively long lifetime, $F(O_3)$ is nearly equal to the net O_3 tendency, $P(O_3)$, in polluted environments. Through this sequence of proportionalities, it is assumed that time-dependent ambient CH_2O concentrations are proportional to the near-instantaneous rate of O_3 production ($P(O_3)$). Plots of these relationships (i.e. CH_2O mixing ratios vs $F(CH_2O)$, $F(CH_2O)$ vs $F(O_3)$,

etc.) are shown in Figures S8-S11 in the supporting information accompanying this work. This derived relationship between CH_2O concentration and $P(O_3)$ is shown in equation XI, where the proportionality factor β is a function of local NO_x concentrations, actinic flux, the respective CH_2O and O_3 branching ratios per hydrocarbon oxidation (α_{CH_2O} and α_{O_3}), and the local lifetime of CH_2O (τ_{CH_2O}). A plot showing the relationship between β, NO, and CH_2O lifetime for data collected in Houston and Maryland is shown in Figure S12 in the supporting information accompanying this work, and β is taken to be the slope of the total least squares fit to these data ($\beta = 1.7 \pm 0.01$ hrs^{-1}), shown in Figure S13.

$$\frac{dO_3}{dt} = \beta(NO_x, J, \alpha_{CH_2O}, \alpha_{O_3}, \tau_{CH_2O})CH_2O(t) \quad \text{(XI)}$$

Since the goal of this section is to create a very broad mathematical understanding of the O_3-CH_2O relationships reported here, β will be simplified and essentially treated as a constant with a value of 1.7 ± 0.01 hrs^{-1}. Integrating equation XI, equation XII is derived:

$$\Delta O_3 = O_{3,t} - O_{3,initial} = \beta \int_{t=0}^{t} CH_2O(t)dt \quad \text{(XII)}$$

In Maryland, a linear fit is appropriate for describing the diurnal trends in CH_2O (see Figure 8). Assuming that CH_2O varies linearly throughout the day with a slope of κ and a y-intercept of $CH_2O_{initial}$ (i.e. equation XIII), results in equation XIV.

$$CH_2O(t) = \kappa t + CH_2O_{initial} \quad \text{(XIII)}$$

$$\Delta O_3 = O_{3,t} - O_{3,initial} = \beta[(CH_2O_{initial})t + \left(\frac{\kappa}{2}\right)t^2] \quad \text{(XIV)}$$

Here, $(CH_2O_{initial})t$ represents the O_3 production associated with the initial lump of CH_2O (i.e. the rate of hydrocarbon oxidation associated with the initial atmospheric composition), and $\kappa t^2/2$ represents additional O_3 production (or loss) associated with increases (or decreases) in CH_2O (i.e. O_3 production or loss associated with deviations from the initial atmospheric composition). Calculating the relative changes in O_3 and CH_2O (i.e. equation XIV divided by equation XIII) under these conditions, results in equation XV:

$$\frac{\Delta O_3}{\Delta CH_2O} = \frac{\beta[(CH_2O_{initial}) + \left(\frac{\kappa}{2}\right)t]}{\kappa} \quad \text{(XV)}$$

On average in Maryland, CH_2O column densities increased by 0.066 DU/hr. Using the average midday CH_2O column density from Maryland (0.61 ± 0.30 DU), $\kappa = 0.066$ DU/hr, and a β value

of 1.7 ± 0.01 hrs^{-1}, results in an average slope of 18.2 ± 0.5 DU O_3/DU CH_2O . This is in close agreement with the slope observed in Maryland in Figure 6.

In Houston, on the other hand, CH_2O tended to remain constant throughout the day because of the dominance of anthropogenic emissions. This results in:

$$O_{3,t} = O_{3,initial} + (\beta CH_2O)t \quad \text{(XVI)}$$

From this equation, it can be seen that, for a given range of CH_2O values (each of which remains constant over time), the slope between O_3 and CH_2O will increase over time. This was simulated in Figure 11 using a range of CH_2O column densities from 0.2 – 1.2 DU, an initial O_3 column density of 9 DU, a β value of 1.7 hrs^{-1}, and times binned every two hours over a ten-hour period. Figure 12 simulates the spatial behavior seen in Houston on September 25th (i.e. Figure 7).

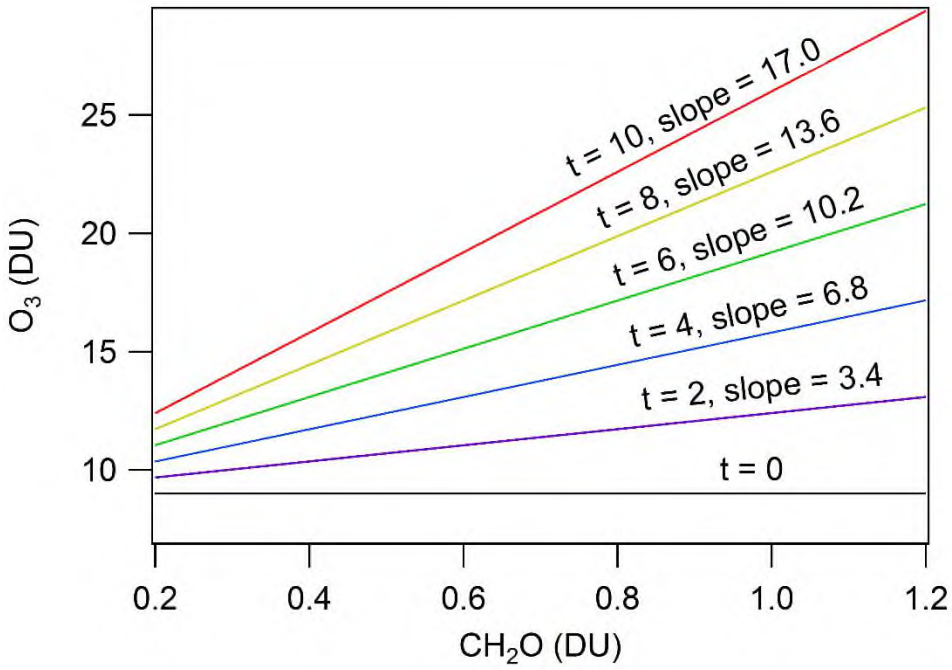


Figure 11. Simulation of equation XVI using a CH_2O range of 0.2 – 1.2 DU, an initial O_3 of 9 DU, and a β value of 1.7.

Deviations from the idealized behavior described in equations XV (for systems in which CH_2O varies linearly throughout the day) and XVI (for systems in which CH_2O remains relatively constant) can come in many forms. In both cases, the proportionality factor β can vary from the average value used here. In the morning and evening, when the CH_2O lifetime is ~3-4 hours and NO_x emissions are relatively high, β is reduced (model estimates predict ~0.5 hrs^{-1}) and a shallower slope would be observed. During midday hours (~10:00 – 14:00) when the CH_2O lifetime is ~1-2 hours and NO_x emissions are relatively low, β would be higher (model

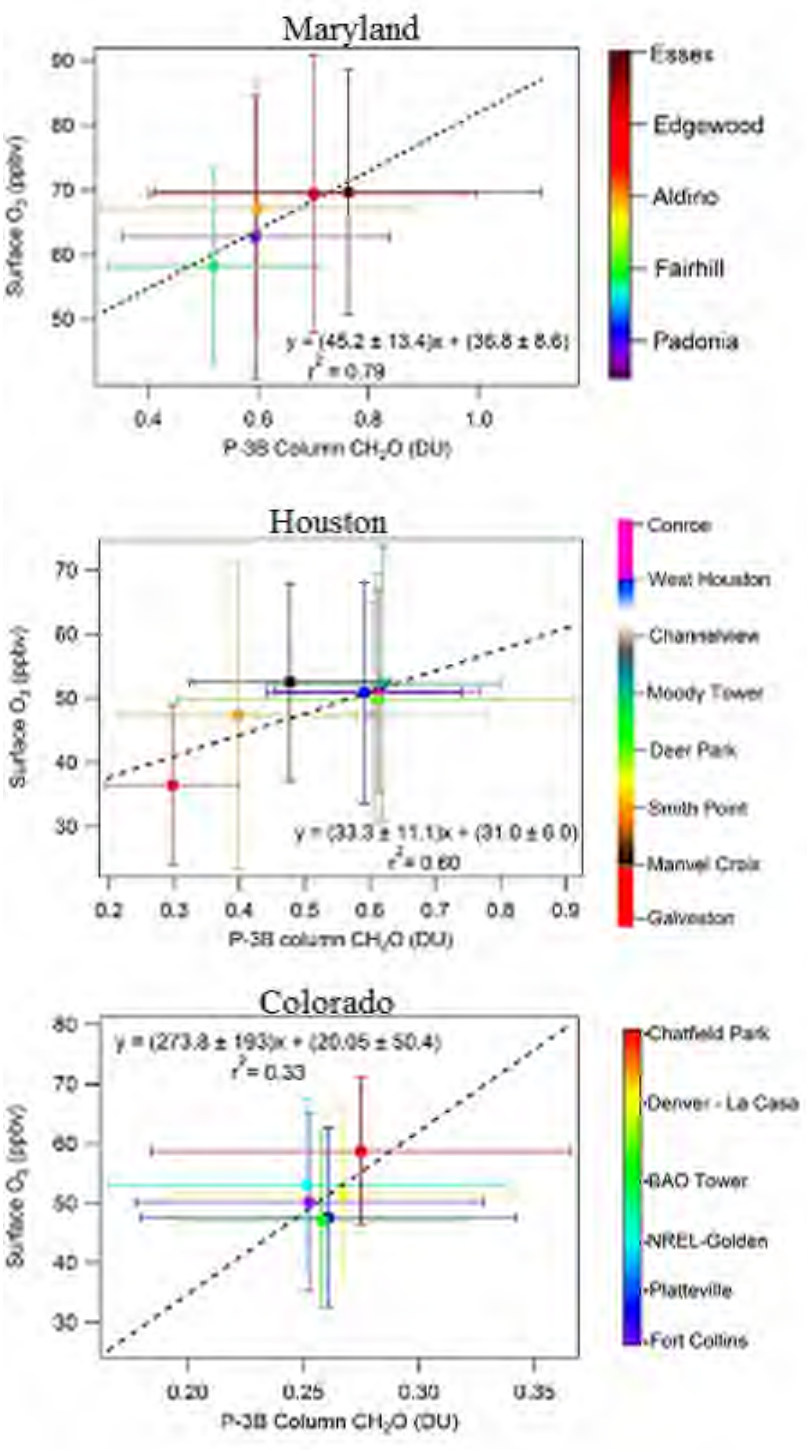
906 estimates predict ~2 hrs^{-1}). Furthermore, transport is not accounted for in these equations, and
907 could be a source of uncertainty (for example, the spatial behavior noted in Houston in section
908 4). However, these equations are not meant to be used as a robust model, but merely provide a
909 mathematical framework for understanding the results presented here.

910 **5.3 Long-Term Averages of Surface O_3 and Column CH_2O**

911 In the previous sections, it was shown to be possible to estimate near-surface O_3 tendencies
912 and gain an understanding of the spatial and temporal trends in near-surface O_3 production from
913 column measurements of CH_2O, and the accuracy of this estimation is much greater when
914 regional emissions are dominated by biogenic VOCs (i.e. Maryland) as opposed to
915 anthropogenic VOCs (i.e. Houston). However, from a human health perspective, estimating O_3
916 exposures at the surface would be much more germane. Previous work has investigated
917 relationships between surface and column measurements of O_3 from DISCOVER-AQ data and
918 found that the relationship can be complex and highly variable in space and time [*Martins et al.*,
919 2013; *Reed et al.*, 2013; *Flynn et al.*, 2014; *Thompson et al.*, 2014]. Furthermore, local
920 meteorological conditions can create stratifications within the lower troposphere, making it very
921 challenging to derive actual surface mixing ratios even from PBL column densities. For example,
922 He et al., (2014) noted a persistent layer with 120 ppbv of O_3 at 800 m from July 18 to 23 during
923 the Maryland deployment of DISCOVER-AQ, while surface O_3 mixing ratios remained near 80
924 ppbv over this period. For these reasons, the previous sections of this work focused on the
925 relationship between column CH_2O and PBL-column O_3 and made no attempt to draw
926 conclusions about concentrations of O_3 at the surface. Factors affecting the O_3 surface/column
927 ratio include transport (both on local and regional scales), meteorology (both on local and
928 regional scales), and temporal and spatial variability in emissions. While there is significant
929 variability in the O_3 surface/column ratio on short time scales, this variability may be muted
930 when averaged over longer timescales. In this section, spatial trends in the O_3-CH_2O relationship
931 are investigated when viewed over a timescale of each DISCOVER-AQ deployment (about one
932 month).

933 Figure 12 shows the relationship between long-term (over the course of each deployment)
934 averages of surface O_3 and whole-column CH_2O in Maryland, Houston, and Colorado, colored
935 by site location. In Maryland and Houston, correlations are sufficient such that long-term O_3
936 exposures could be reasonably estimated from long-term averaging of column measurements of
937 CH_2O at discrete locations. While the variability bars for both CH_2O column densities and
938 surface O_3 mixing ratios in Figure 12 are quite large (representing a wide range of day-to-day
939 conditions), the mean values for each location have reasonable correlations in Maryland and
940 Houston. In Colorado, however, the dynamic range in CH_2O is substantially smaller than in
941 either Maryland or Houston, and the resulting correlation is weak. In Houston, five sites (Conroe,
942 West Houston, Moody Tower, Channelview, and Deer Park) have similar long-term average
943 values of CH_2O column density and surface O_3 mixing ratios. As a result, the slope and r^2 values
944 for the Houston data likely do not represent the true values that would be seen if locations were
945 chosen to maximize the dynamic range in surface O_3 and column CH_2O without producing
946 clusters of similar numbers. While this highlights one of the shortcomings of the DISCOVER-
947 AQ dataset (we are limited to data collected at the spiral sites – 8 places in Houston, 5 in
948 Maryland, etc.), it also provides an interesting glimpse into how future air quality monitoring
949 networks could be optimized. Using a mobile network of co-located surface measurements of O_3
950 and surface-based column measurements of CH_2O (from a Pandora spectrometer, for example), a

more representative relationship between long-term averages of column CH_2O and surface O_3 could be created by optimizing measurement sites to maximize the dynamic ranges of the measurements. Then, this relationship could be applied to long-term averages of satellite measurements of CH_2O at discrete locations to map out long-term exposures to surface O_3 in between surface monitoring sites – effectively providing a map of estimated long-term O_3 exposure throughout an entire region.



957

958 **Figure 12.** Long-term (over the course of each deployment) averages of surface O_3 versus
959 whole-column CH_2O in Maryland (top), Houston (middle), and Colorado (bottom). Error bars
960 represent the 1σ variability for each site. Note that each graph has its own unique range for both
961 the x and y axes.

962

6 Conclusion

Using data collected during NASA's DISCOVER-AQ field campaigns, a correlation between column measurements of CH_2O and O_3 was observed. When the O_3 column was restricted to only include O_3 contained within the chemical PBL, the correlation improved for data collected in Maryland, but had no effect on data collected in Houston. Further analysis of these data revealed that O_3 and CH_2O were more strongly correlated when temporal changes in CH_2O were present. By creating a filter that removed data that had little temporal variation in CH_2O, a strong correlation emerged ($r^2 = 0.75$) between O_3 and CH_2O in Maryland. In Houston, however, there was no improvement to the weak correlation between O_3 and CH_2O. A case study of September 25th in Houston – the day that had the highest surface O_3 mixing ratios observed in Houston – revealed the reason for the poor correlation: By viewing the data spatially rather than temporally, relatively strong correlations emerged and suggested that a time-of-transport term for plumes originating from the ship channel was the cause of the weak overall correlation between O_3 and CH_2O in Houston.

Further analysis provided an explanation for the different behaviors that were observed between Maryland and Houston. In Maryland, biogenic emissions fueled diurnal increases in VOCR, which in turn caused CH_2O to, on average, increase monotonically throughout the day. Because O_3 has a much longer atmospheric lifetime and accumulates throughout the day, CH_2O and O_3 co-varied in Maryland. In Houston, however, anthropogenic emissions from the ship channel dominated the local VOCR. As a result, biogenics played a proportionately small role in local CH_2O production, and no discernable diurnal trend in VOCR or CH_2O was observed. Using the LaRC photochemical box model, we showed that ambient CH_2O mixing ratios change in response to changes in the local hydrocarbon oxidation rate – primarily due to changes in OH or VOC reactivity. Model results suggested that, because O_3 has a much longer lifetime and accumulates throughout the day, a strong correlation between O_3 and CH_2O will emerge when CH_2O mixing ratios increase over time in conjunction with increasing emissions of hydrocarbon precursors. Finally, it was found that long-term averages of CH_2O (i.e. weekly or monthly averages) could be useful for estimating long-term O_3 exposure at the surface.

While the results of this work provide insights into the potential utility of future geostationary satellites to fill in gaps in surface monitoring networks and understand the true spatial extent of air pollution events, these results are unlikely to be applicable globally. As discussed above, the O_3-CH_2O relationship is most useful in regions where temporal variability in VOC emissions drive changes in ambient CH_2O mixing ratios throughout the day (i.e. Maryland). In regions where CH_2O levels are low (as was the case in the California and Colorado deployments of DISCOVER-AQ), satellites may not be able to sufficiently resolve gradients and this relationship cannot be applied at all. Furthermore, NO_x has a complex, non-linear effect on the chemical processes that produce/destroy O_3 and CH_2O, and therefore can affect the slope of the relationship between column O_3 and CH_2O. Because emission rates of biogenic hydrocarbons and NO_x can vary greatly both regionally and seasonally, studies invoking the use of 3D models would be invaluable in determining regions and seasons where the O_3-CH_2O relationship would be most applicable. Guided by such model results, locations could be identified for siting co-located surface measurements and surface-based column measurements (such as Pandora spectrometers or tethered balloons) of O_3, CH_2O, and NO_2. Such measurements would be invaluable for further evaluating these results and determining the

locations and conditions for which the covariance of surface ozone and column CH_2O would be a useful metric from satellite observations.

7 Acknowledgements

All data used in this work can be downloaded from: http://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html. PTR-MS measurements during DISCOVER-AQ were supported by the Austrian Ministry for Transport, Innovation, and Technology through the Austrian Space Applications Program (ASAP) of the Austrian Research Promotion Agency (FFG). Jason Schroeder and Tomas Mikoviny were partially supported by an appointment with the NASA Postdoctoral Program at NASA Langley Research Center, administered by Oak Ridge Associated Universities through a contract with NASA. Armin Wisthaler and Markus Müller received support from the Visiting Scientist Program of the National Institute of Aerospace (NIA). The authors would like to thank the thoughtful reviewers who gave valuable input into making this paper better.

8 References

Anderson, B. E., G. L. Gregory, J. E. Collins, G. W. Sachse, T. J. Conway, and G. P. Whiting (1996), Airborne observations of spatial and temporal variability of tropospheric carbon dioxide, *J. Geophys. Res.*, *101*(D1), 1985–1997, doi:10.1029/95JD00413.

Apel, E. C. et al. (2012), Impact of the deep convection of isoprene and other reactive trace species on radicals and ozone in the upper troposphere, *Atmos. Chem. Phys.*, *12*(2), 1135–1150, doi:10.5194/acp-12-1135-2012.

Baidar, S., H. Oetjen, S. Coburn, B. Dix, I. Ortega, R. Sinreich, and R. Volkamer (2013), The CU Airborne MAX-DOAS instrument: Vertical profiling of aerosol extinction and trace gases, *Atmos. Meas. Tech.*, *6*(3), 719–739, doi:10.5194/amt-6-719-2013.

Boersma, K. F., D. J. Jacob, H. J. Eskes, R. W. Pinder, J. Wang, and R. J. van der A (2008), Intercomparison of SCIAMACHY and OMI tropospheric NO_2 columns: Observing the diurnal evolution of chemistry and emissions from space, *J. Geophys. Res. Atmos.*, *113*(16), 1–14, doi:10.1029/2007JD008816.

Boersma, K. F., D. J. Jacob, M. Trainic, Y. Rudich, I. DeSmedt, R. Dirksen, and H. J. Eskes (2009), Validation of urban NO_2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities, *Atmos. Chem. Phys.*, *9*(12), 3867–3879, doi:10.5194/acp-9-3867-2009.

Colman, J., A. Swanson, S. Meinardi, D. R. Sive B., Blake, and F. S. Rowland (2001), Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Anal. Chem.*, (73), 3723–3731.

Cooper, O. R. et al. (2014), Global distribution and trends of tropospheric ozone: An observation-based review, *Elem. Sci. Anthr.*, *2*, 000029, doi:10.12952/journal.elementa.000029.

Crawford, J. et al. (1999), Assessment of upper tropospheric HO_x sources over the tropical

Pacific based on NASA GTE/PEM data: Net effect on HO_x and other photochemical parameters, *J. Geophys. Res. Atmos.*, *104*(D13), 16225–16273.

Diskin, G. S., J. R. Podolske, G. W. Sachse, and T. A. Slate (2002), Open-Path Airborne Tunable Diode Laser Hygrometer, in *Proceedings of the Society for Photo-Optical Instrumentation Engineers (SPIE)*, vol. 4817, edited by A. Fried, pp. 196–204.

Downey, N., C. Emery, J. Jung, T. Sakulyanontvittaya, L. Hebert, D. Blewitt, and G. Yarwood (2015), Emission reductions and urban ozone responses under more stringent US standards, *Atmos. Environ.*, *101*, 209–216, doi:10.1016/j.atmosenv.2014.11.018.

Duncan, B. N. et al. (2010), Application of OMI observations to a space-based indicator of NO_x and VOC controls on surface ozone formation, *Atmos. Environ.*, *44*(18), 2213–2223, doi:10.1016/j.atmosenv.2010.03.010.

Duncan, B. N. et al. (2014), Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid, *Atmos. Environ.*, *94*, 647–662, doi:10.1016/j.atmosenv.2014.05.061.

Fishman, J. et al. (2008), Remote Sensing of Tropospheric Pollution from Space, *Bull. Am. Meteorol. Soc.*, *89*(6), 805–821, doi:10.1175/2008BAMS2526.1.

Flynn, C. M. et al. (2014), Relationship between column-density and surface mixing ratio: Statistical analysis of O_3 and NO_2 data from the July 2011 Maryland DISCOVER-AQ mission, *Atmos. Environ.*, *92*, 429–441, doi:10.1016/j.atmosenv.2014.04.041.

Fried, A. et al. (2008), Role of convection in redistributing formaldehyde to the upper troposphere over North America and the North Atlantic during the summer 2004 INTEX campaign, *J. Geophys. Res.*, *113*(D17), D17306, doi:10.1029/2007JD009760.

Fried, A. et al. (2011), Detailed comparisons of airborne formaldehyde measurements with box models during the 2006 INTEX-B and MILAGRO campaigns: Potential evidence for significant impacts of unmeasured and multi-generation volatile organic carbon compounds, *Atmos. Chem. Phys.*, *11*(22), 11867–11894, doi:10.5194/acp-11-11867-2011.

Fried, A., C. P. Loughner, and K. Pickering (2016), *Analysis of Airborne Formaldehyde Data Over Houston Texas Acquired During the 2013 DISCOVER-AQ and SEAC4RS Campaigns*. Report prepared for Texas AQRP.

Garcia, A. R., R. Volkamer, L. T. Molina, M. J. Molina, J. Samuelson, J. Mellqvist, B. Galle, S. C. Herndon, and C. E. Kolb (2006), Separation of emitted and photochemical formaldehyde in Mexico City using a statistical analysis and a new pair of gas-phase tracers, *Atmos. Chem. Phys.*, *6*, 4545–4557, doi:10.5194/acp-6-4545-2006.

Geron, C., P. Harley, and A. Guenther (2001), Isoprene emission capacity for US tree species, *Atmos. Environ.*, *35*(19), 3341–3352, doi:10.1016/S1352-2310(00)00407-6.

Hache, E. et al. (2014), The added value of a visible channel to a geostationary thermal infrared instrument to monitor ozone for air quality, *Atmos. Meas. Tech.*, *7*(7), 2185–2201, doi:10.5194/amt-7-2185-2014.

He, H. et al. (2014), An elevated reservoir of air pollutants over the Mid-Atlantic States during the 2011 DISCOVER-AQ campaign: Airborne measurements and numerical simulations, *Atmos. Environ.*, *85*, 18–30, doi:10.1016/j.atmosenv.2013.11.039.

Hudman, R. C. et al. (2004), Ozone production in transpacific Asian pollution plumes and implications for ozone air quality in California, *J. Geophys. Res. D Atmos.*, *109*, 1–14, doi:10.1029/2004JD004974.

Ichoku, C., R. Kahn, and M. Chin (2012), Satellite contributions to the quantitative characterization of biomass burning for climate modeling, *Atmos. Res.*, *111*, 1–28, doi:10.1016/j.atmosres.2012.03.007.

Jin, X., and T. Holloway (2015), Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument, *J. Geophys. Res. Atmos.*, *120*, 7229–7246, doi:10.1002/2015JD023250

Junkermann, W. (2009), On the distribution of formaldehyde in the western Po-Valley, Italy, during FORMAT 2002/2003, *Atmos. Chem. Phys. Discuss.*, *9*(3), 13999–14022, doi:10.5194/acpd-9-13999-2009.

Kefauver, S. C., I. Filella, and J. Peñuelas (2014), Remote sensing of atmospheric biogenic volatile organic compounds (BVOCs) via satellite-based formaldehyde vertical column assessments, *Int. J. Remote Sens.*, *35*(21), 7519–7542, doi:10.1080/01431161.2014.968690.

Kleinman, L. I. (1994), Low and high NOx tropospheric photochemistry, *J. Geophys. Res.*, *99*(94), 831–838.

Lee, Y. et al. (1998), Atmospheric chemistry and distribution of formaldehyde and several multioxygenated carbonyl compounds during the 1995 Nashville/Middle Tennessee Ozone Study, *J. Geophys. Res.*, *103*, 22449 – 22462, doi:10.1029/98JD01251.

Li, S.-M. L., K. G. Anlauf, H. A. Wiebe, and J. W. Bottenheim (1994), Estimating primary and secondary production of HCHO in eastern North America based on gas phase measurements and principal component analysis among the components species against the absolute principal component scores, *Geo*, *21*(8), 669–672.

Lin, M., A. M. Fiore, O. R. Cooper, L. W. Horowitz, A. O. Langford, H. Levy, B. J. Johnson, V. Naik, S. J. Oltmans, and C. J. Senff (2012a), Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions, *J. Geophys. Res. Atmos.*, *117*(D21), doi:10.1029/2012JD018151.

Lin, Y. C., J. J. Schwab, K. L. Demerjian, M.-S. Bae, W.-N. Chen, Y. Sun, Q. Zhang, H.-M. Hung, and J. Perry (2012b), Summertime formaldehyde observations in New York City: Ambient levels, sources and its contribution to HO_x radicals, *J. Geophys. Res.*, *117*(D8), D08305, doi:10.1029/2011JD016504.

Loughner, C. P. et al. (2014), Impact of bay-breeze circulations on surface air quality and boundary layer export, *J. Appl. Meteorol. Climatol.*, *53*(7), 1697–1713, doi:10.1175/JAMC-D-13-0323.1.

Macdonald, A. M., P. A. Makar, K. G. Anlauf, K. L. Hayden, J. W. Bottenheim, D. Wang, and T. Dann (2001), Summertime formaldehyde at a high-elevation site in Quebec at the site .

Data were segregated based on the ratio of NO_x to NO_y and on the inversely related to the NO_x / NO_y ratio . During the high HCHO episode, *J. Geophys. Res.*, *106*, 32361 – 32374, doi:10.1029/2001JD000513.

Mahajan, A. S., I. De Smedt, M. S. Biswas, S. Ghude, S. Fadnavis, C. Roy, and M. van Roozendael (2015), Inter-annual variations in satellite observations of nitrogen dioxide and formaldehyde over India, *Atmos. Environ.*, *116*, 194–201, doi:10.1016/j.atmosenv.2015.06.004.

Marais, E. A. et al. (2012), Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns, *Atmos. Chem. Phys.*, *12*(14), 6219–6235, doi:10.5194/acp-12-6219-2012.

Martin, R. V. (2008), Satellite remote sensing of surface air quality, *Atmos. Environ.*, *42*(34), 7823–7843, doi:10.1016/j.atmosenv.2008.07.018.

Martin, R. V., D. D. Parrish, T. B. Ryerson, J. K. Nicks, K. Chance, T. P. Kurosu, D. J. Jacob, E. D. Sturges, A. Fried, and B. P. Wert (2004), Evaluation of GOME satellite measurements of tropospheric NO_2 and HCHO using regional data from aircraft campaigns in the southeastern United States, *J. Geophys. Res. D Atmos.*, *109*(24), 1–11, doi:10.1029/2004JD004869.

Martins, D. K., R. M. Stauffer, A. M. Thompson, H. S. Halliday, D. Kollonige, E. Joseph, and A. J. Weinheimer (2013), Ozone correlations between mid-tropospheric partial columns and the near-surface at two mid-atlantic sites during the DISCOVER-AQ campaign in July 2011, *J. Atmos. Chem.*, *72*(3-4), 373–391, doi:10.1007/s10874-013-9259-4.

Millet, D. B. et al. (2006), Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission, *J. Geophys. Res. Atmos.*, *111*(24), 1–17, doi:10.1029/2005JD006853.

Millet, D. B., D. J. Jacob, K. F. Boersma, T.-M. Fu, T. P. Kurosu, K. Chance, C. L. Heald, and A. Guenther (2008), Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, *J. Geophys. Res.*, *113*(D2), D02307, doi:10.1029/2007JD008950.

Mukund, R., T. J. Kelly, and C. W. Spicer (1996), Source attribution of ambient air toxic and other VOCs in Columbus, Ohio, *Atmos. Environ.*, *30*(20), 3457–3470, doi:10.1016/1352-2310(95)00487-4.

Muller, M. et al. (2014), A compact PTR-ToF-MS instrument for airborne measurements of volatile organic compounds at high spatiotemporal resolution, *Atmos. Meas. Tech.*, *7*(11), 3763–3772, doi:10.5194/amt-7-3763-2014.

Murphy, C. F., and D. T. Allen (2005), Hydrocarbon emissions from industrial release events in the Houston-Galveston area and their impact on ozone formation, *Atmos. Environ.*, *39*(21), 3785–3798, doi:10.1016/j.atmosenv.2005.02.051.

National Science and Technology Council (2013), *Air Quality Observation Systems in the United States, Product of the Committee on Environment, Natural Resources, and Sustainability*.

Natraj, V. et al. (2011), Multi-spectral sensitivity studies for the retrieval of tropospheric and

lowermost tropospheric ozone from simulated clear-sky GEO-CAPE measurements, *Atmos. Environ.*, *45*(39), 7151–7165, doi:10.1016/j.atmosenv.2011.09.014.

Olson, J. R. et al. (2001), Seasonal differences in the photochemistry of the South Pacific: A comparison of observations and model results from PEM-Tropics A and B, *J. Geophys. Res.*, *106*(D23), 32749–32766, doi:10.1029/2001JD900077.

Olson, J. R., J. H. Crawford, G. Chen, W. H. Brune, I. C. Faloona, D. Tan, H. Harder, and M. Martinez (2006), A reevaluation of airborne HO_x observations from NASA field campaigns, *J. Geophys. Res. Atmos.*, *111*(10), 1–12, doi:10.1029/2005JD006617.

Oltmans, S. J. et al. (2006), Long-term changes in tropospheric ozone, *Atmos. Environ.*, *40*(17), 3156–3173, doi:10.1016/j.atmosenv.2006.01.029.

Palmer, P. I., D. J. Jacob, A. M. Fiore, R. V. Martin, K. Chance, and T. P. Kurosu (2003), Mapping isoprene emissions over North America using formaldehyde column observations from space, *J. Geophys. Res.*, *108*(D6), 4180, doi:10.1029/2002JD002153.

Palmer, P. I. et al. (2006), Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, *J. Geophys. Res.*, *111*(D12), D12315, doi:10.1029/2005JD006689.

Paoletti, E., A. De Marco, D. C. S. Beddows, R. M. Harrison, and W. J. Manning (2014), Ozone levels in European and USA cities are increasing more than at rural sites, while peak values are decreasing., *Environ. Pollut.*, *192*(x), 295–9, doi:10.1016/j.envpol.2014.04.040.

Parrish, D. D. et al. (2012), Primary and secondary sources of formaldehyde in urban atmospheres: Houston Texas region, *Atmos. Chem. Phys.*, *12*(7), 3273–3288, doi:10.5194/acp-12-3273-2012.

Paulot, F., J. D. Crounse, H. G. Kjaergaard, J. H. Kroll, J. H. Seinfeld, and P. O. Wennberg (2009), Isoprene photooxidation: new insights into the production of acids and organic nitrates, *Atmos. Chem. Phys.*, *9*(4), 1479–1501, doi:10.5194/acp-9-1479-2009.

Pfister, G. G., L. K. Emmons, P. G. Hess, J. F. Lamarque, J. J. Orlando, S. Walters, A. Guenther, P. I. Palmer, and P. J. Lawrence (2008), Contribution of isoprene to chemical budgets: A model tracer study with the NCAR CTM MOZART-4, *J. Geophys. Res. Atmos.*, *113*(5), 1–21, doi:10.1029/2007JD008948.

Potosnak, M. J., L. LeStourgeon, S. G. Pallardy, K. P. Hosman, L. Gu, T. Karl, C. Geron, and A. B. Guenther (2014), Observed and modeled ecosystem isoprene fluxes from an oak-dominated temperate forest and the influence of drought stress, *Atmos. Environ.*, *84*, 314–322, doi:10.1016/j.atmosenv.2013.11.055.

Reed, A. J., A. M. Thompson, D. E. Kollonige, D. K. Martins, M. A. Tzortziou, J. R. Herman, T. A. Berkoff, N. K. Abuhassan, and A. Cede (2013), Effects of local meteorology and aerosols on ozone and nitrogen dioxide retrievals from OMI and pandora spectrometers in Maryland, USA during DISCOVER-AQ 2011, *J. Atmos. Chem.*, *72*(3-4), 455–482, doi:10.1007/s10874-013-9254-9.

Richter, D., P. Weibring, J. G. Walega, A. Fried, S. M. Spuler, and M. S. Taubman (2015), Compact highly sensitive multi-species airborne mid-IR spectrometer, *Appl. Phys. B Lasers*

Opt., 119–131, doi:10.1007/s00340-015-6038-8.

Ryerson, T. B. (2003), On tropospheric ozone formation in Houston, Texas, *J. Geophys. Res.*, *108*(x), 1–24, doi:10.1029/2002JD003070.

Sachse, G. W., J. E. Collins, G. F. Hill, L. O. Wade, L. G. Burney, and J. A. Ritter (1991), Airborne tunable diode laser sensor for high-precision concentration and flux measurements of carbon monoxide and methane, in *Proceedings of the Society for Photo-Optical Instrumentation Engineers (SPIE)*, vol. 1433, edited by H. I. Schiff, pp. 157–166.

Schroeder, J. R., L. L. Pan, T. Ryerson, G. Diskin, J. Hair, S. Meinardi, I. Simpson, B. Barletta, N. Blake, and D. R. Blake (2014), Evidence of mixing between polluted convective outflow and stratospheric air in the upper troposphere during DC3, *J. Geophys. Res. Atmos.*, *119*(19), 11477–11491, doi:10.1002/2014JD022109.

Sillman, S., J. Logan, and S. Wofsy (1990), The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes, *J. Geophys.*, *95*, 1837–1851, doi:10.1029/JD095iD02p01837.

Sillman, S. et al. (1995), Photochemistry of Ozone Formation in Atlanta,GA-Models and Measurements, *Atmos. Environ.*, *29*(21), 3055–3066.

Simpson, I. J. et al. (2010), Characterization of trace gases measured over Alberta oil sands mining operations: 76 speciated C_2–C_{10} volatile organic compounds (VOCs), CO_2, CH_4, CO, NO, NO_2, NO_y, O_3, *Atmos. Chem. Phys.*, *10*(23), 11931–11954, doi:10.5194/acp-10-11931-2010.

Simpson, I. J., M. P. S. Andersen, S. Meinardi, L. Bruhwiler, N. J. Blake, D. Helmig, F. S. Rowland, and D. R. Blake (2012), Long-term decline of global atmospheric ethane concentrations and implications for methane, *Nature*, *488*(7412), 490–494, doi:10.1038/nature11342.

De Smedt, I. et al. (2015), Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, *Atmos. Chem. Phys.*, *15*(21), 12519–12545, doi:10.5194/acp-15-12519-2015.

Streets, D. G. et al. (2013), Emissions estimation from satellite retrievals: A review of current capability, *Atmos. Environ.*, *77*, 1011–1042, doi:10.1016/j.atmosenv.2013.05.051.

Sumner, A. L. et al. (2001), A study of formaldehyde chemistry above a forest canopy, *J. Geophys. Res.*, *106*(D20), 24387, doi:10.1029/2000JD900761.

Taraborrelli, D., M. G. Lawrence, T. M. Butler, R. Sander, and J. Lelieveld (2008), Mainz Isoprene Mechanism 2 (MIM2): an isoprene oxidation mechanism for regional and global atmospheric modelling, *Atmos. Chem. Phys. Discuss.*, *8*(4), 14033–14085, doi:10.5194/acpd-8-14033-2008.

Thompson, A. M., R. M. Stauffer, S. K. Miller, D. K. Martins, E. Joseph, A. J. Weinheimer, and G. S. Diskin (2014), Ozone profiles in the Baltimore-Washington region (2006-2011): Satellite comparisons and DISCOVER-AQ observations, *J. Atmos. Chem.*, *72*(3-4), 393–422, doi:10.1007/s10874-014-9283-z.

US EPA (2015), *National Ambient Air Quality Standards for Ozone*.

Valin, L. C., A. R. Russell, and R. C. Cohen (2014), Chemical feedback effects on the spatial patterns of the NO_x weekend effect: a sensitivity analysis, *Atmos. Chem. Phys.*, *14*(1), 1–9, doi:10.5194/acp-14-1-2014.

Valin, L. C., M. Fiore, A, K. Chance, and G. González Abad (2015), The role of OH production in interpreting the variability of CH_2O columns in the southeast U.S., *J. Geophys. Res. Atmos.*, *121*, 3510–3532, doi:10.1002/2015JD024012.

Vay, S. A., B. E. Anderson, T. J. Conway, G. W. Sachse, J. E. Collins, D. R. Blake, and D. J. Westberg (1999), Airborne observations of the tropospheric CO_2 distribution and its controlling factors over the South Pacific Basin, *J. Geophys. Res.*, *104*(D5), 5663–5676, doi:10.1029/98JD01420.

Vizuete, W., B. U. Kim, H. Jeffries, Y. Kimura, D. T. Allen, M. A. Kioumourtzoglou, L. Biton, and B. Henderson (2008), Modeling ozone formation from industrial emission events in Houston, Texas, *Atmos. Environ.*, *42*(33), 7641–7650, doi:10.1016/j.atmosenv.2008.05.063.

Weibring, P., D. Richter, J. G. Walega, L. Rippe, and A. Fried (2010), Difference frequency generation spectrometer for simultaneous multispecies detection., *Opt. Express*, *18*(26), 27670–27681, doi:10.1364/OE.18.027670.

Weinheimer, A. J., J. G. Walega, B. A. Ridley, B. L. Gary, D. R. Blake, N. J. Blake, F. S. Rowland, G. W. Sachse, B. E. Anderson, and J. E. Collins (1994), Meridional distributions of NO_X, NO_y, and other species in the lower statosphere and upper troposphere during AASE 2, *Geophys. Res. Let.*, *21*(23), 2583–2586.

Wert, B. P. (2003), Signatures of terminal alkene oxidation in airborne formaldehyde measurements during TexAQS 2000, *J. Geophys. Res.*, *108*(D3), 4104, doi:10.1029/2002JD002502.

Wolfe, G. M. et al. (2015), Formaldehyde production from isoprene oxidation across NOx regimes, *Atmos. Chem. Phys. Discuss.*, *15*(21), 31587–31620, doi:10.5194/acpd-15-31587-2015.

Wolfe, G. M. et al. (2016), Formaldehyde production from isoprene oxidation across NOx regimes, *Atmos. Chem. Phys.*, *16*(4), 2597–2610, doi:10.5194/acp-16-2597-2016.

Zhang, L., D. J. Jacob, X. Yue, N. V. Downey, D. a. Wood, and D. Blewitt (2014), Sources contributing to background surface ozone in the US Intermountain West, *Atmos. Chem. Phys.*, *14*(11), 5295–5309, doi:10.5194/acp-14-5295-2014.

Zhu, L., D. J. Jacob, L. J. Mickley, E. A. Marais, D. S. Cohan, Y. Yoshida, B. N. Duncan, G. González Abad, and K. V Chance (2014), Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns, *Environ. Res. Lett.*, *9*(11), 114004, doi:10.1088/1748-9326/9/11/114004.

Zoogman, P., D. J. Jacob, K. Chance, L. Zhang, P. Le Sager, A. M. Fiore, A. Eldering, X. Liu, V. Natraj, and S. S. Kulawik (2011), Ozone air quality measurement requirements for a geostationary satellite mission, *Atmos. Environ.*, *45*(39), 7143–7150,

doi:10.1016/j.atmosenv.2011.05.058.

Zoogman, P., D. J. Jacob, K. Chance, X. Liu, M. Lin, a. Fiore, and K. Travis (2014), Monitoring high-ozone events in the US Intermountain West using TEMPO geostationary satellite observations, *Atmos. Chem. Phys.*, *14*(12), 6261–6271, doi:10.5194/acp-14-6261-2014.

Zoogman, P., X. Liu, K. Chance, Q. Sun, C. Schaaf, T. Mahr, and T. Wagner (2016), A climatology of visible surface reflectance spectra, *J. Quant. Spectrosc. Radiat. Transf.*, *180*, 39–46, doi:10.1016/j.jqsrt.2016.04.003.

Figure 9.

Manvel Croix 13:10

Galveston 13:30

Smith Point 13:45

Moody Tower 14:10

West Houston 14:20

Conroe 14:45

Channelview, 15:10

Deer Park 15:20

CH_2O (measured)

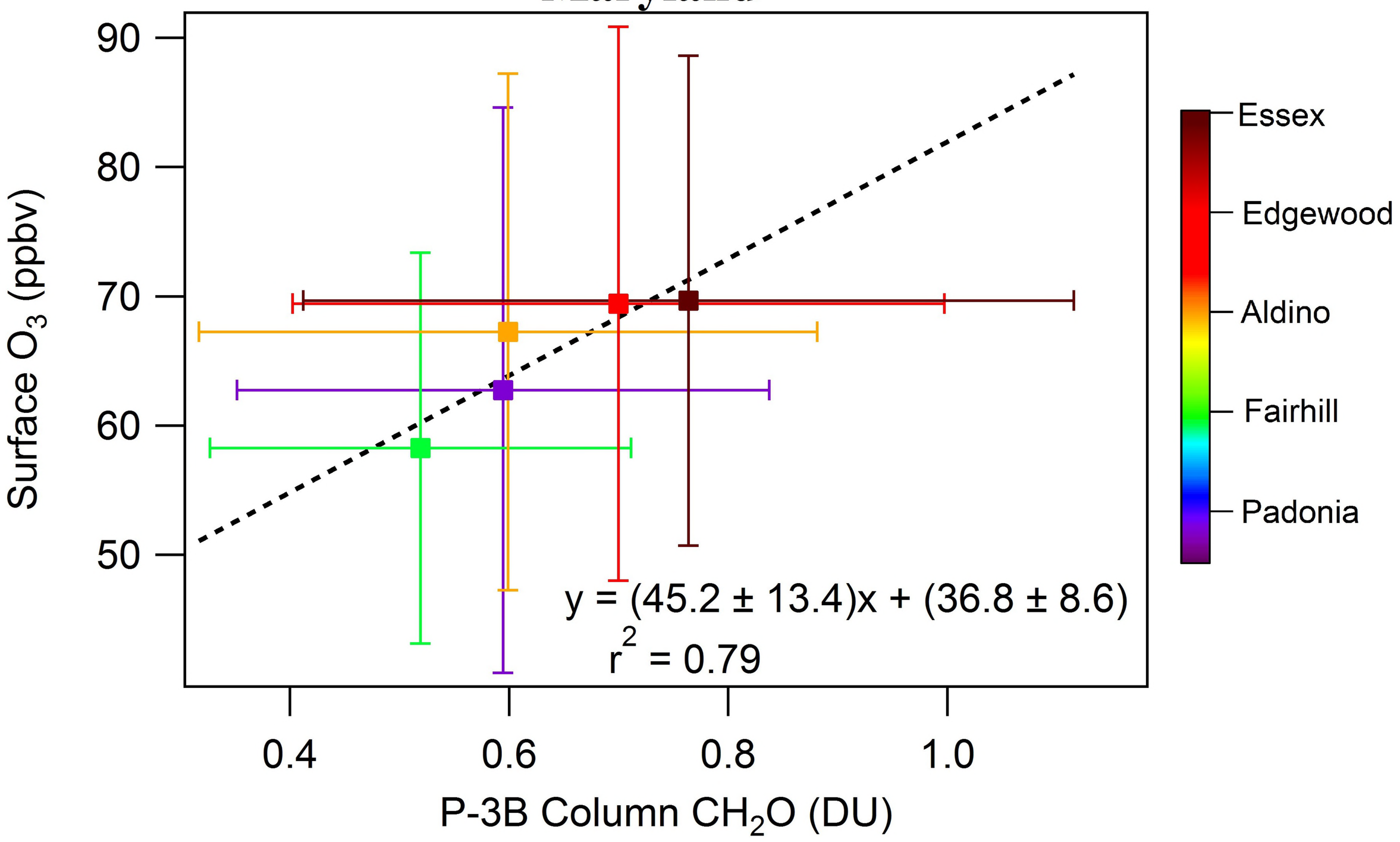
CH_2O (model)

Altitude (km)

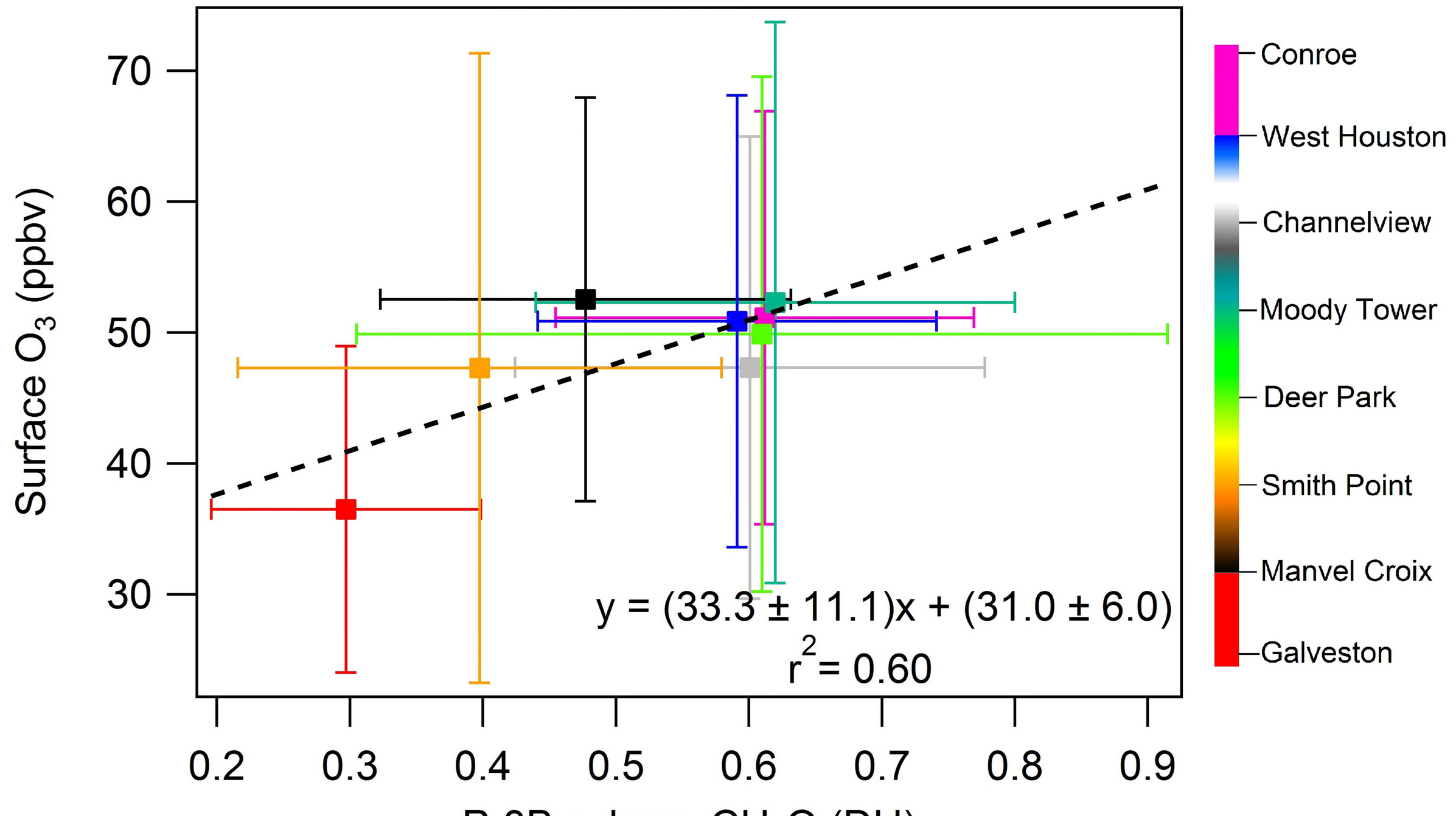
CH_2O (pptv)

Figure 13.

Maryland



Houston



Colorado

Figure 12.

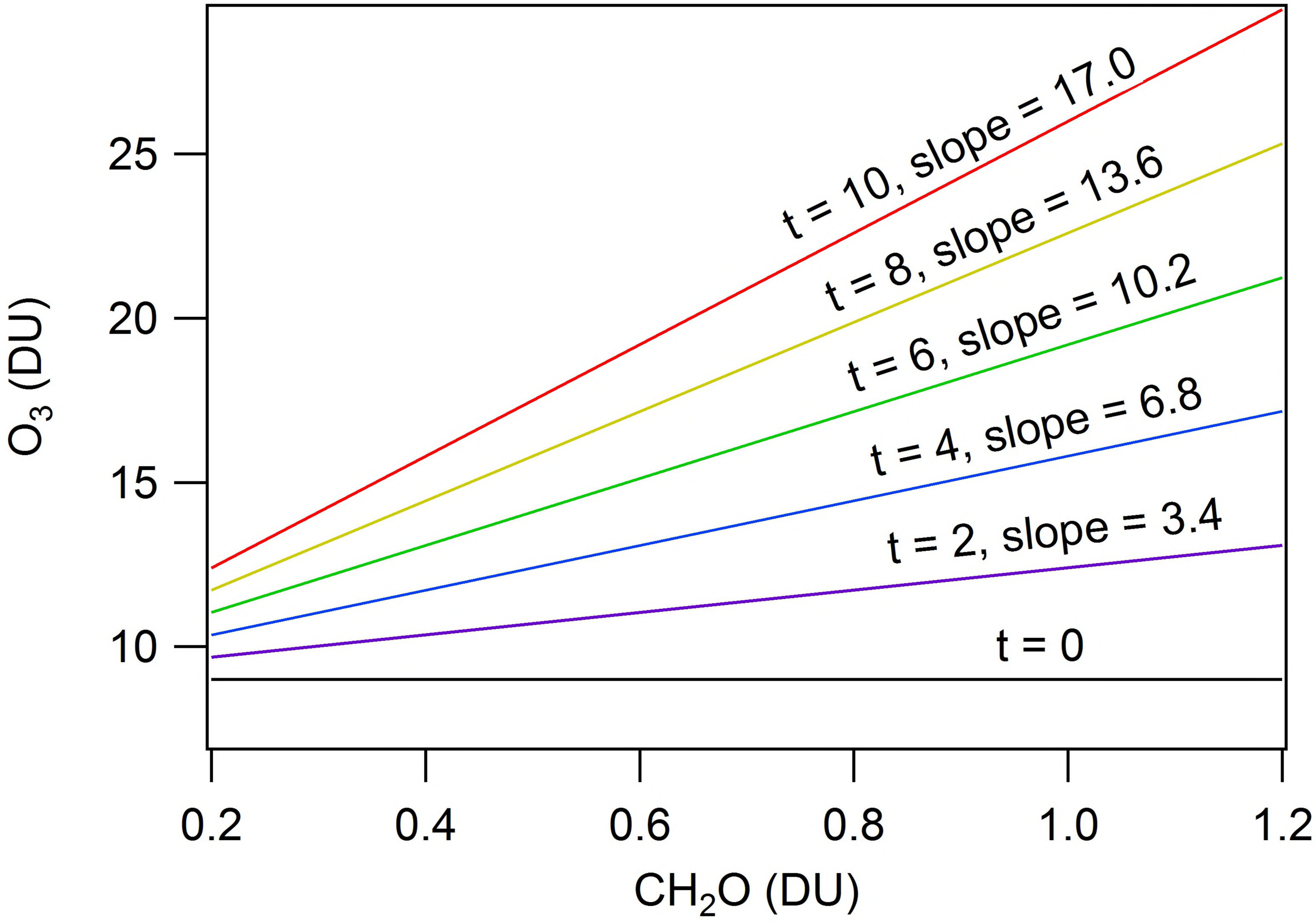

t = 10, slope = 17.0
t = 8, slope = 13.6
t = 6, slope = 10.2
t = 4, slope = 6.8
t = 2, slope = 3.4
t = 0
25
20
15
10
0.2
0.4
0.6
0.8
1.0
1.2
O3 (DU)
CH2O (DU)

Figure 11.

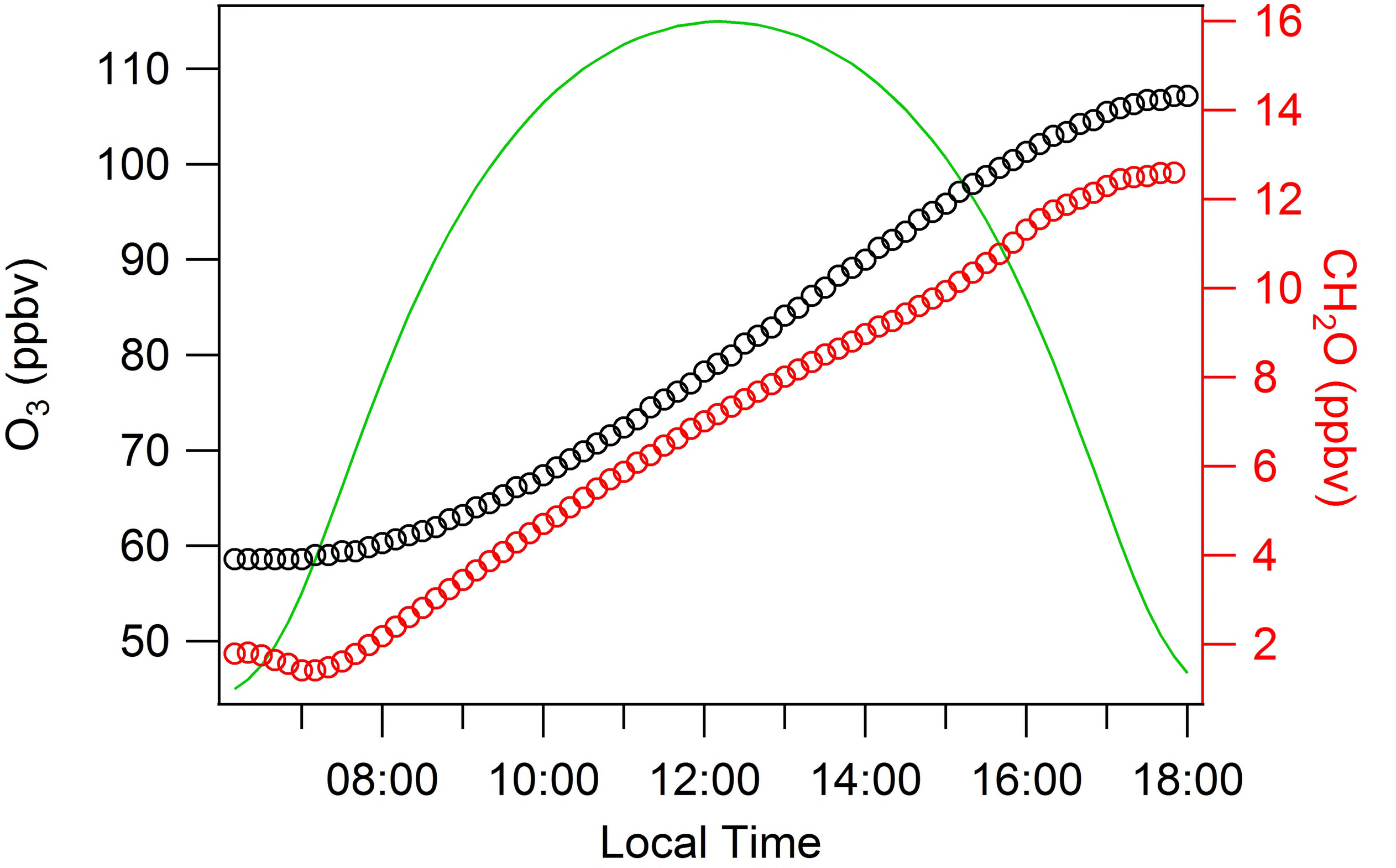
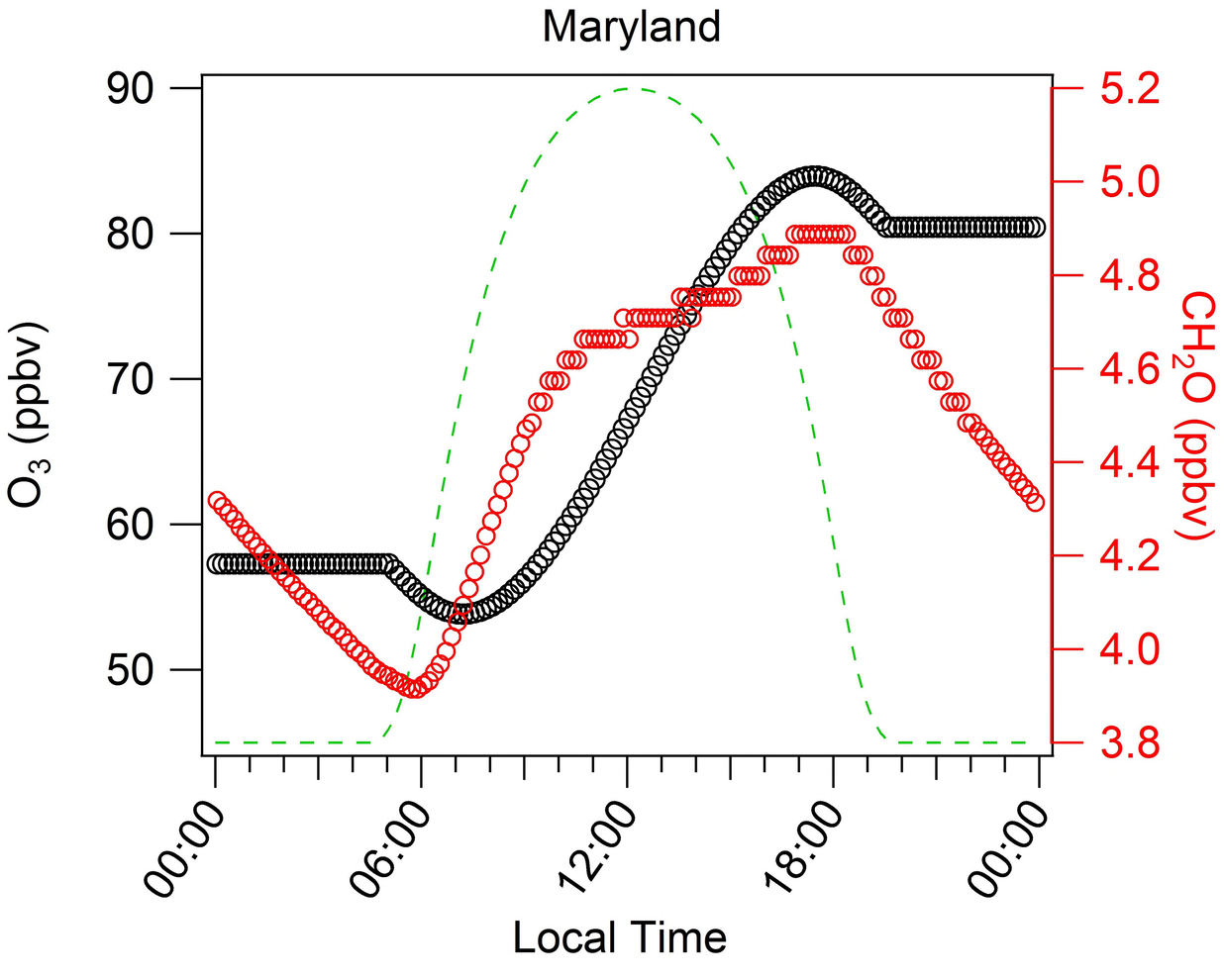

110
100
90
80
70
60
50
O_3 (ppbv)
16
14
12
10
8
6
4
2
CH_2O (ppbv)
08:00
10:00
12:00
14:00
16:00
18:00
Local Time

Figure 10.


Maryland
90
80
70
60
50
O_3 (ppbv)
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
CH_2O (ppbv)
00:00
06:00
12:00
18:00
00:00
Local Time

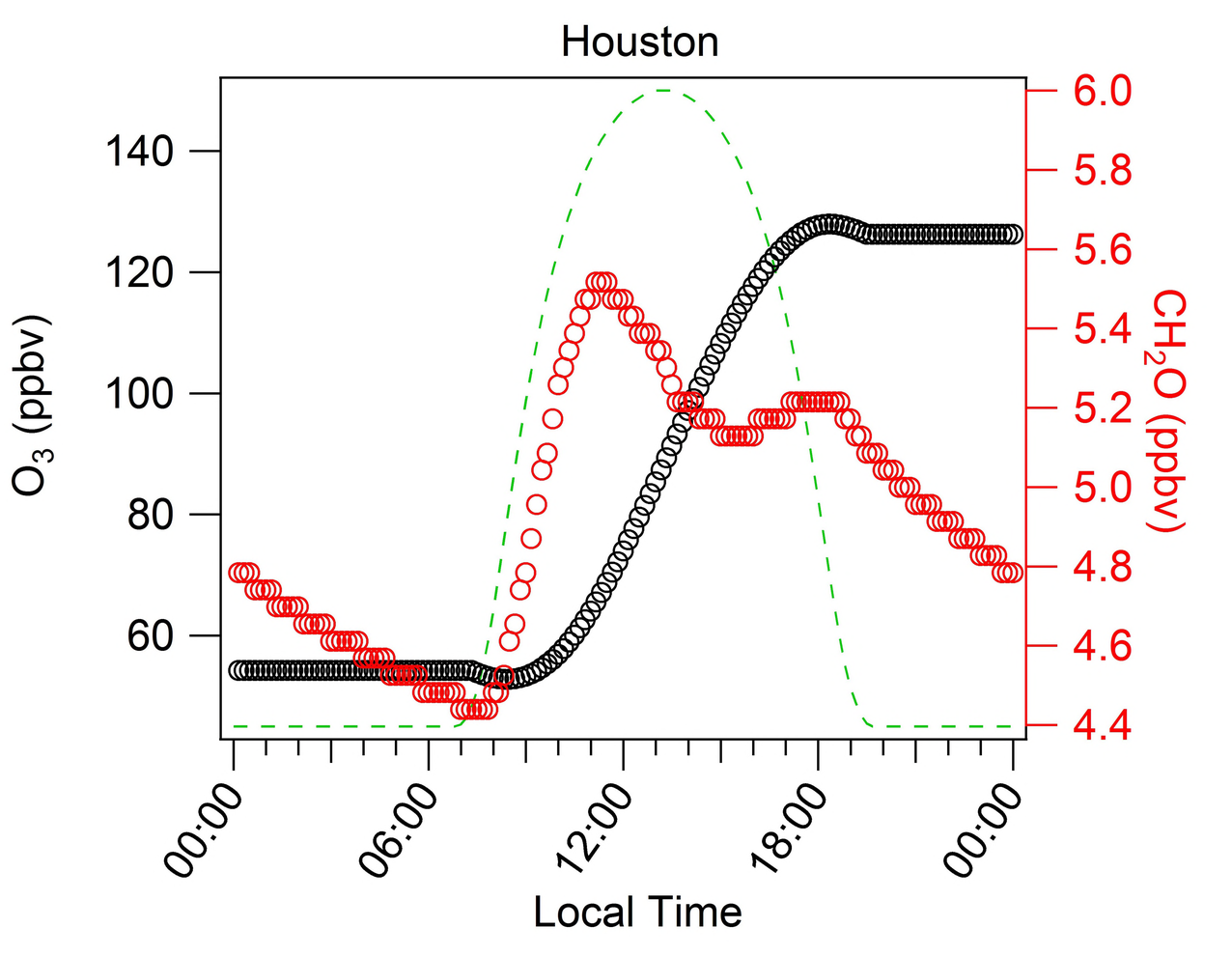

Houston
140
120
100
80
60
O_3 (ppbv)
6.0
5.8
5.6
5.4
5.2
5.0
4.8
4.6
4.4
CH_2O (ppbv)
00:00
06:00
12:00
18:00
00:00
Local Time

Figure 8.

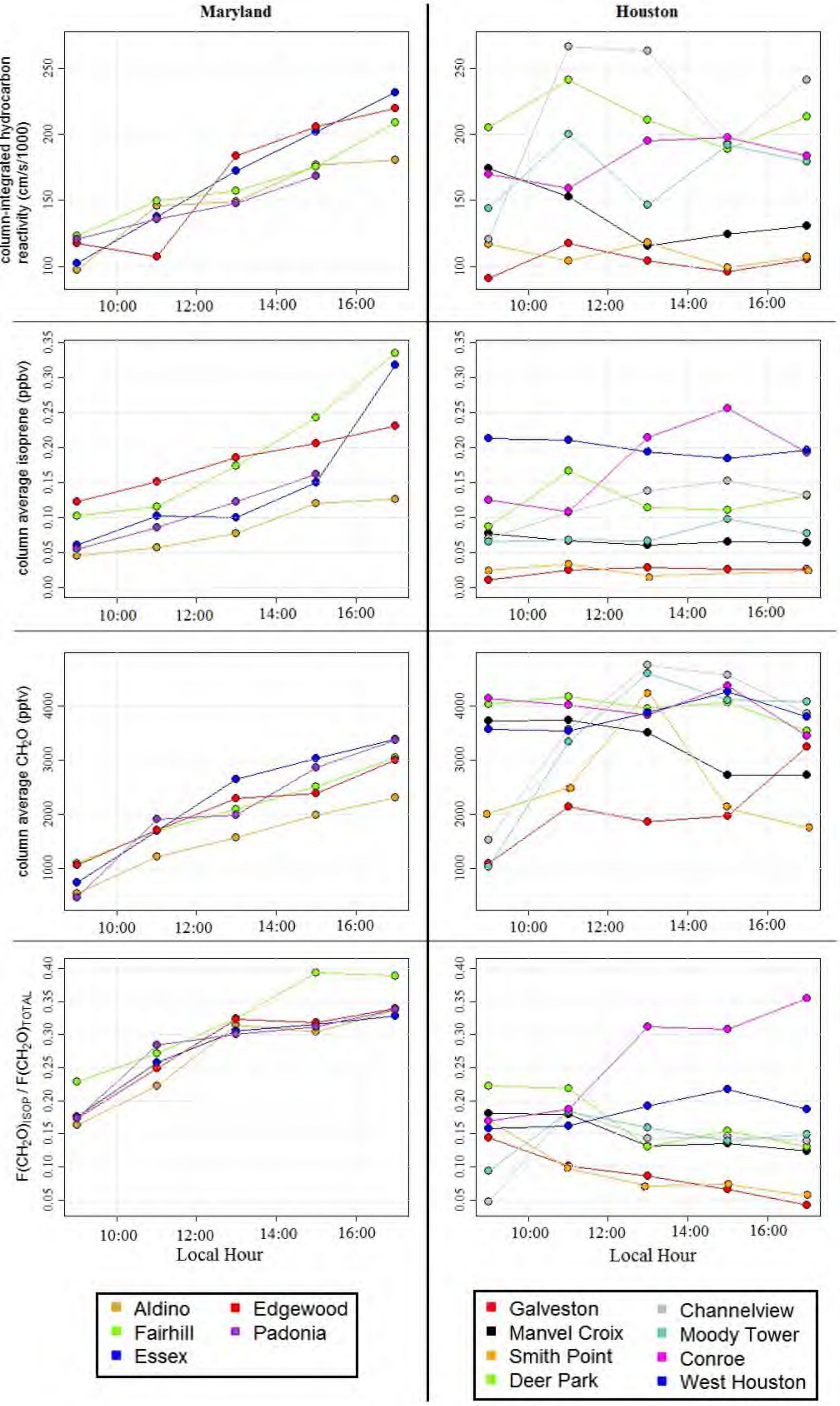

Maryland
Houston
column-integrated hydrocarbon reactivity (cm/s/1000)
column average isoprene (ppbv)
column average CH_2O (pptv)
$F(CH_2O)_{ISOP} / F(CH_2O)_{TOTAL}$
10:00
12:00
14:00
16:00
Local Hour
Aldino
Edgewood
Fairhill
Padonia
Essex
Galveston
Channelview
Manvel Croix
Moody Tower
Smith Point
Conroe
Deer Park
West Houston

Figure 7.

Houston, Sept 25

y = 11.6x + 8.7

r^2 = 0.86

y = 8.6x + 7.3

r^2 = 0.31

y = 8.1x + 7.3

r^2 = 0.66

20

15

10

$O_{3,\ PBL}$ (DU)

0.4

0.6

0.8

1.0

1.2

1.4

CH_2O (DU)

Channelview

Conroe

Deer Park

Galveston

Manvel Croix

Moody Tower

Smith Point

West Houston

Midday fit

Afternoon fit

Whole day fit

Figure 6.

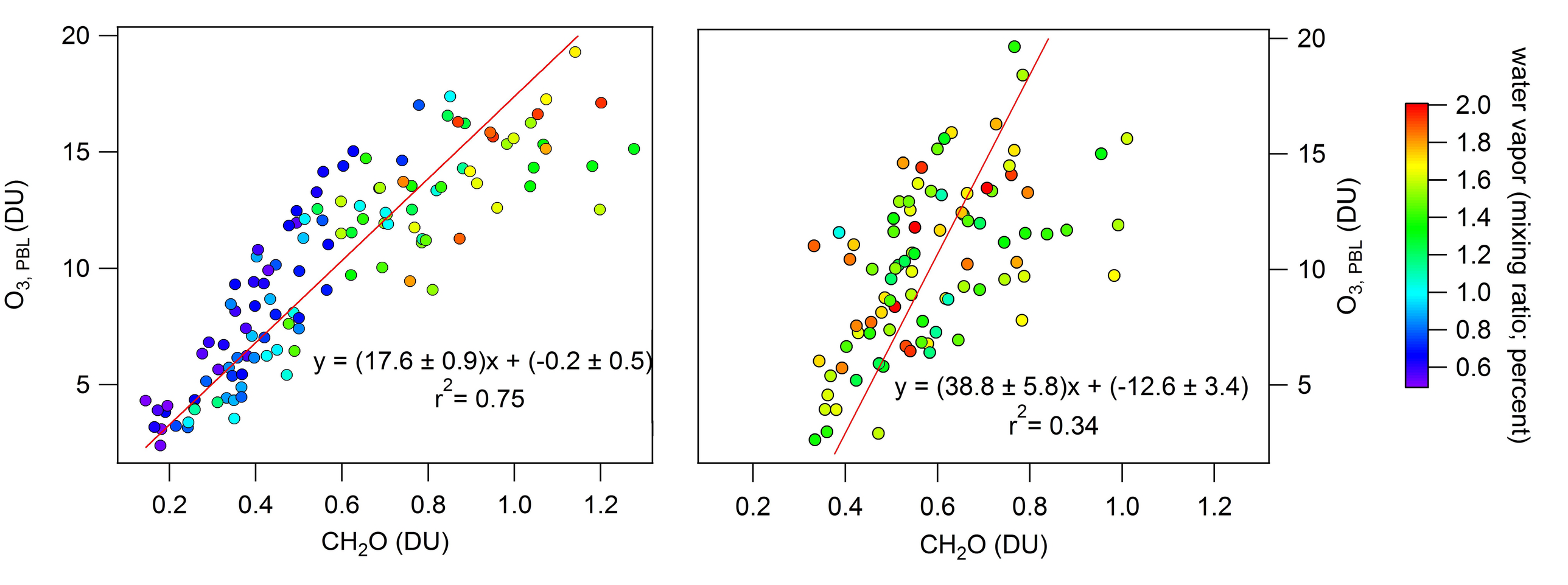

20
15
10
5
$O_{3,\ PBL}$ (DU)
y = (17.6 ± 0.9)x + (-0.2 ± 0.5)
r^2= 0.75
0.2
0.4
0.6
0.8
1.0
1.2
CH_2O (DU)
y = (38.8 ± 5.8)x + (-12.6 ± 3.4)
r^2= 0.34
water vapor (mixing ratio; percent)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6

Figure 5.

July 26
July 27
July 28
July 29

$O_{3,\ PBL}$ (DU)

15
10
5
0

$y = (19.1 \pm 1.2)x + (-1.3 \pm 0.8)$

$r^2 = 0.87$

0.0
0.2
0.4
0.6
0.8
1.0
1.2

CH_2O (DU)

Figure 4.

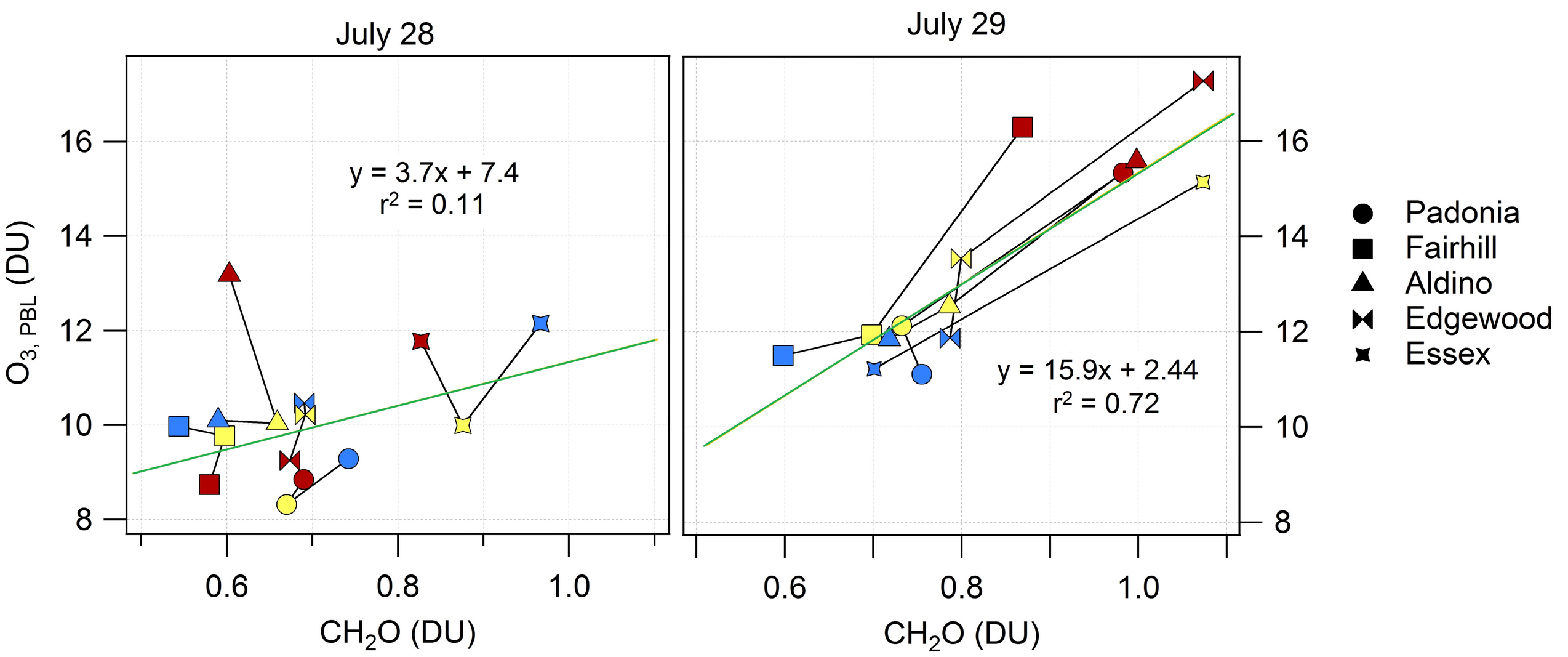

July 28
July 29
y = 3.7x + 7.4
r2 = 0.11
y = 15.9x + 2.44
r2 = 0.72
O3, PBL (DU)
CH2O (DU)
CH2O (DU)
8
10
12
14
16
0.6
0.8
1.0
Padonia
Fairhill
Aldino
Edgewood
Essex

Figure 3.

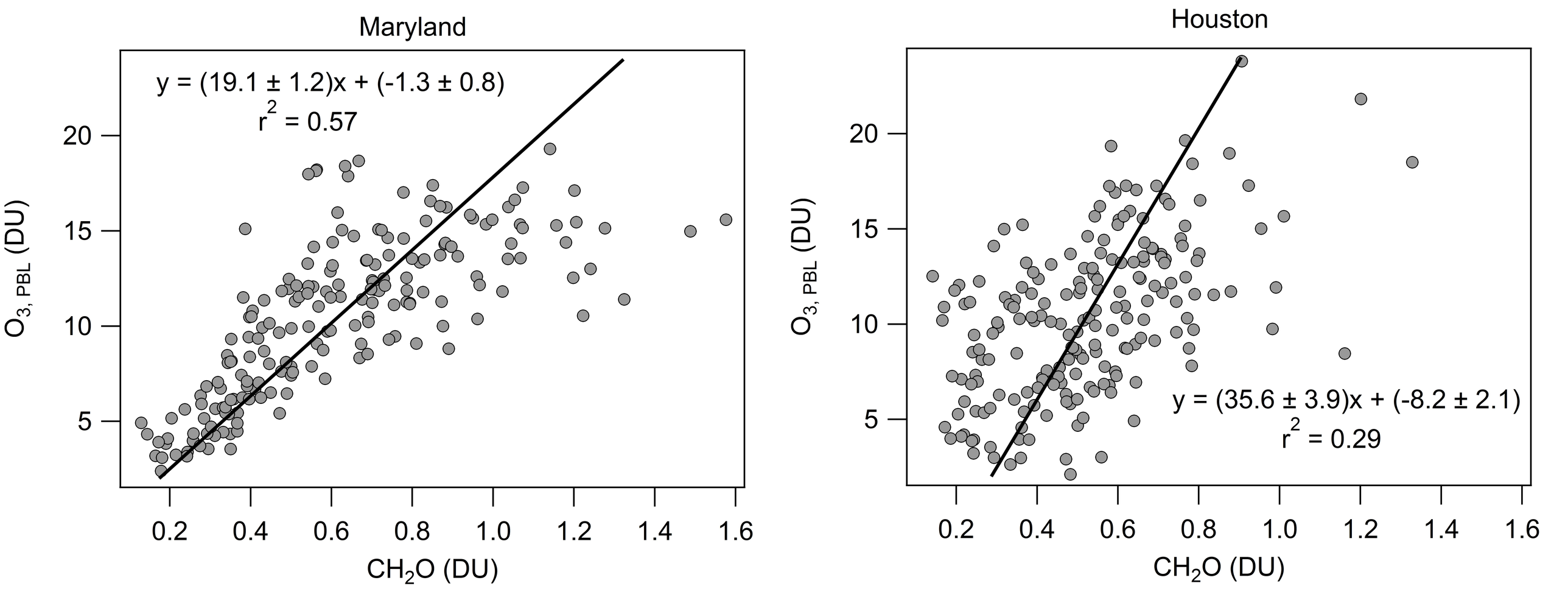

Maryland
y = (19.1 ± 1.2)x + (-1.3 ± 0.8)
r2 = 0.57
O3, PBL (DU)
CH2O (DU)
Houston
y = (35.6 ± 3.9)x + (-8.2 ± 2.1)
r2 = 0.29
O3, PBL (DU)
CH2O (DU)

Figure 2.

Maryland (July 2011)
California (Jan 2013)
Houston (Sept 2013)
Colorado (Aug 2014)

O_3 (DU)

30
25
20
15
10
5
0

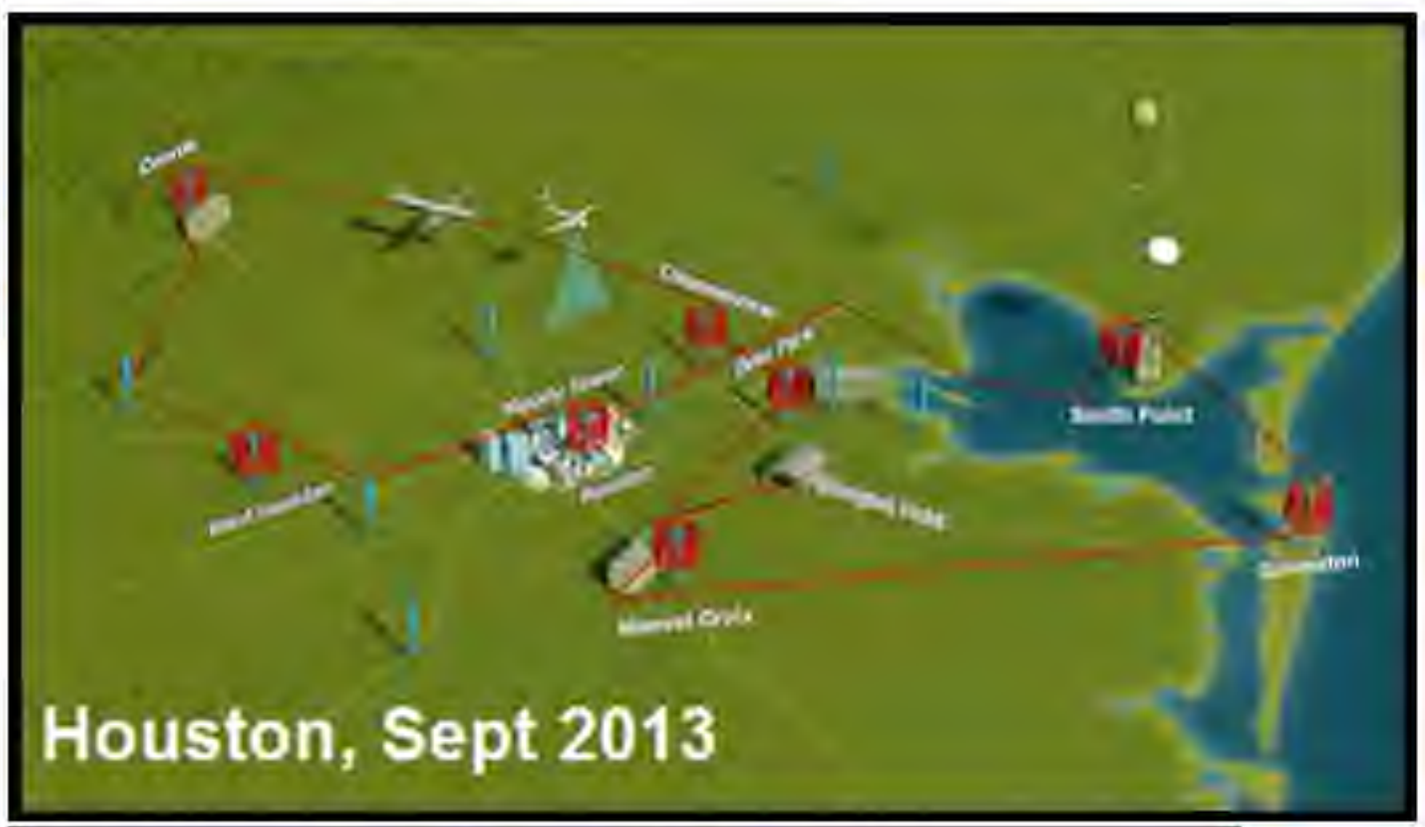
y = 19.3x + 6.5
r^2 = 0.45

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6

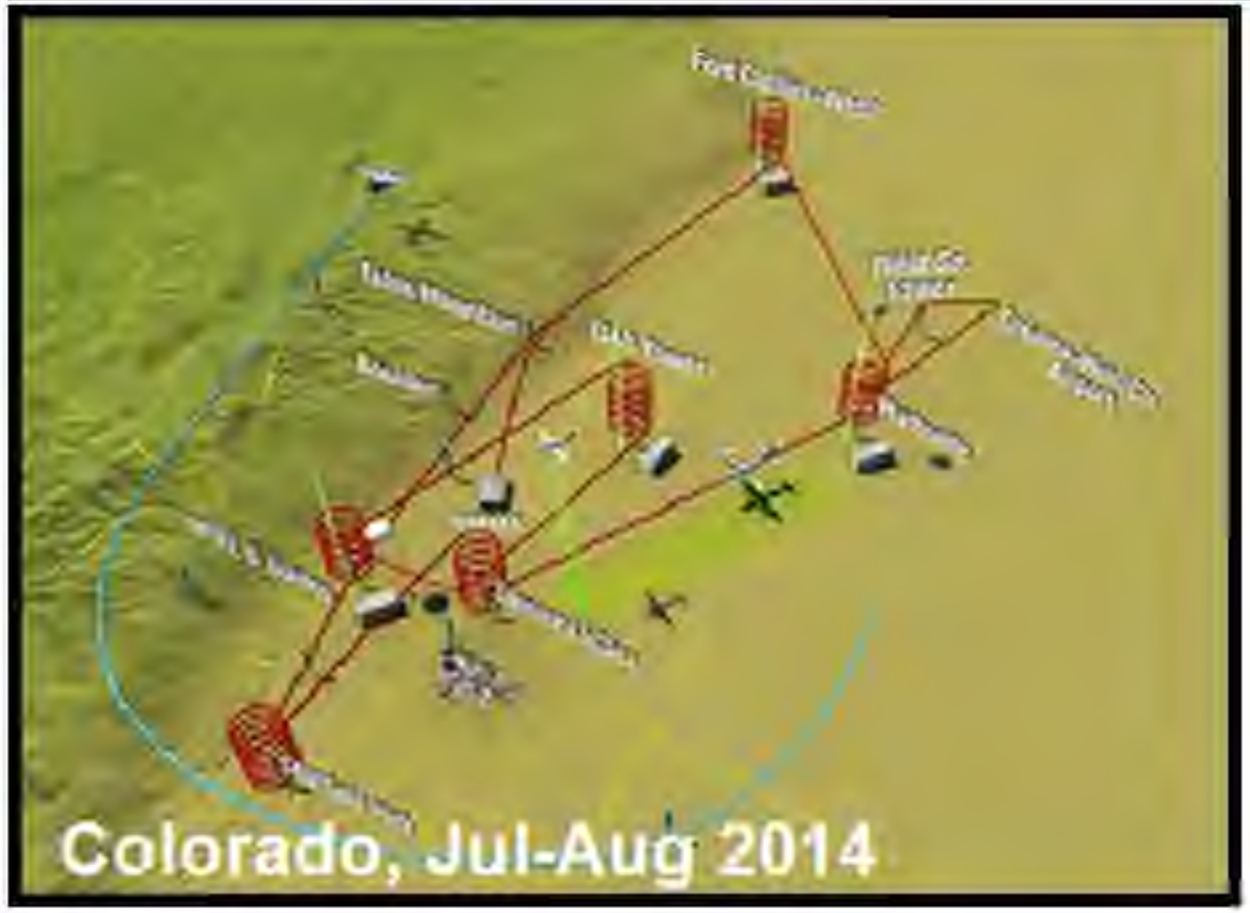
CH_2O (DU)

Figure 1.


Maryland, July 2011


Houston, Sept 2013


California, Jan-Feb 2013


Colorado, Jul-Aug 2014